Random Linear Estimation With Rotationally-Invariant Designs: Asymptotics at High Temperature

Yufan Li, Zhou Fan, Subhabrata Sen, and Yihong Wu

***Abstract*— We study estimation in the linear model $y = A\beta^\star + \epsilon$, in a Bayesian setting where $\beta^\star$ has an entrywise i.i.d. prior and the design A is rotationally-invariant in law. In the large system limit as dimension and sample size increase proportionally, a set of related conjectures have been postulated for the asymptotic mutual information, Bayes-optimal mean squared error, and TAP mean-field equations that characterize the Bayes posterior mean of $\beta^\star$. In this work, we prove these conjectures for a general class of signal priors and for arbitrary rotationally-invariant designs A, under a "high-temperature" condition that restricts the range of eigenvalues of $A^\top A$ and encompasses regimes of sufficiently low signal-to-noise ratio. Our proof uses a conditional second-moment method argument, where we condition on the iterates of a version of the Vector AMP algorithm for solving the TAP mean-field equations.**

***Index Terms*— Estimation, mutual information, multiaccess communication, Bayes methods.**

I. Introduction

CONSIDER observations $y = A\beta^\star + \epsilon \in \mathbb{R}^m$ from a linear model with Gaussian noise, in a Bayesian setting where the entries of $\beta^\star \in \mathbb{R}^n$ are drawn i.i.d. from a "signal prior". Fundamental questions of interest in applications spanning CDMA communication systems [1] to sparse signal recovery [2] to statistical genetics [3] pertain to the properties of the Bayes posterior law and posterior mean estimate for $\beta^\star$.

In the asymptotic limit as $m, n \to \infty$ and A constitutes an i.i.d. measurement design, a rich and insightful body of literature has obtained precise "single-letter" characterizations of the asymptotic mutual information, minimum mean squared error (MMSE), and low-dimensional marginals of the Bayes posterior law. Based initially on work of Tanaka [1] and Guo and Verdú [4] using the non-rigorous replica method of statistical physics, these characterizations have since been proven rigorously in increasingly general contexts [5], [6], [7], [8], [9], [10], [11], and are closely connected to Approximate Message Passing (AMP) algorithms and mean-field variational approaches for Bayesian inference.

The focus of our work, and the subject of significant recent attention, is on extensions of these results beyond i.i.d. designs. We study here the model where A is right-rotationally invariant in law, and where analogous single-letter characterizations are expected to depend on A only via the spectral distribution of $A^\top A$. In this model, conjectures for the asymptotic mutual information were derived via the replica method for binary and Bernoulli-Gaussian signal priors by Takeda, Uda, and Kabashima [12] and Tulino et al. [13], and a form for general priors was stated by Barbier et al. in [14]. A number of iterative Bayesian inference algorithms including Adaptive TAP [15], [16], Expectation-Consistency [17], Vector/Orthogonal AMP [18], [19], [20], [21], and long-memory forms of AMP [22], [23] have been proposed for this model, whose algorithmic fixed-points coincide with the replica predictions. In [24], the forms of the TAP mean-field equations that characterize the posterior mean were derived for this and related models using a Plefka expansion approach, and it was argued that the approximations underlying these AdaTAP, EC, and VAMP/OAMP algorithms are all equivalent to the vanishing of certain diagrammatic terms in the Plefka expansion. Recently, extensions of the replica method calculations and analyses of VAMP fixed points have been carried out in [25] for settings with a possibly mismatched likelihood or Bayesian prior.

In this work, we provide a rigorous proof of the expressions for asymptotic mutual information and MMSE and of the validity of the TAP mean-field equations (in an L^2 sense) that are predicted by this replica theory, in a setting of correctly specified likelihood and prior, for general rotationally-invariant designs A under a restriction for the range of eigenvalues of $A^\top A$. The centered matrix $A^\top A - d_* I$ for a constant $d_* > 0$ plays the role of a rotationally-invariant couplings matrix in analogous models of mean-field spin glasses [26], [27], and our restriction on the eigenvalue range is analogous to an assumption of high temperature in such spin glass models. In the current statistical context, this assumption encompasses regimes of sufficiently low signal-to-noise ratio (SNR). Our results are complementary to those of [14] that established the asymptotic mutual information for specific designs of the form $A = BW$ where W has i.i.d. Gaussian entries, without a high-temperature constraint. Related analyses of convex empirical risk minimization for linear and generalized

Manuscript received 5 January 2023; revised 7 September 2023; accepted 16 September 2023. Date of publication 5 October 2023; date of current version 16 February 2024. An earlier version of this paper was presented in part at the 2023 IEEE International Symposium on Information Theory (ISIT) [DOI: 10.1109/ISIT54713.2023.10206572]. *(Corresponding author: Yufan Li.)*

Yufan Li and Subhabrata Sen are with the Department of Statistics, Harvard University, Cambridge, MA 02138 USA (e-mail: yufan_li@g.harvard.edu; subhabratasen@fas.harvard.edu).

Zhou Fan and Yihong Wu are with the Department of Statistics, Yale University, New Haven, CT 06520 USA (e-mail: zhou.fan@yale.edu; yihong.wu@yale.edu).

Communicated by R. Venkataramanan, Associate Editor for Machine Learning and Statistics, Communications, Signal Processing and Source Coding.

Digital Object Identifier 10.1109/TIT.2023.3321575

linear models with rotationally-invariant designs have also been performed in [28] and [29], where replica predictions for the minimum mean-squared error were rigorously established.

Interest in the linear model with rotationally-invariant measurement designs has been partially motivated by the belief that asymptotic predictions derived for such designs may hold universally across designs whose right singular vectors are sufficiently "generic". Universality statements of this form have been shown recently for AMP and other first-order iterative algorithms in [30], [31], and [32]. These results suggest that the asymptotic mutual information and Bayes optimal MMSE are also potentially universal—we leave this as an open question for future work.

Proof Ideas: Our proofs build upon recent analyses of orthogonally-invariant spin glasses [33], [34] via a conditional second-moment argument [35], [36]. We analyze the first and second moments of restrictions of the log-partition function conditioned on iterates of a version of VAMP, to establish the asymptotic mutual information. The intuition for this proof strategy is common with [33] and [35], in that the leading order contribution to the fluctuation of the log-partition function (conjecturally) depends on the data (y, A) only via the posterior mean $\mathbb{E}[\beta^\star \mid y, A]$, and thus tight bounds for the log-partition function may be obtained via a second-moment analysis conditional on $\mathbb{E}[\beta^\star \mid y, A]$. To enable explicit calculations, we condition instead on a sequence of VAMP iterates that converge to $\mathbb{E}[\beta^\star \mid y, A]$. At a technical level, this extends analyses of [33] to encompass a general class of prior distributions with possibly unbounded support, and to address additional complexities of the Hamiltonian and of the VAMP algorithm for the linear model. Related ideas of analyzing the conditional moments of a truncated log-partition function in an Ising perceptron model with unbounded log-activation were developed recently in [37].

Let us clarify that existing state-evolution analyses of the mean squared error achieved by VAMP imply only an upper bound for the Bayes-optimal MMSE. Integrating the I-MMSE relation [38] from zero SNR to small positive SNR via an "area argument" [5], [6] then implies a corresponding upper bound for the asymptotic mutual information at high temperature. The main contribution of our work is to prove that these upper bounds are tight, and that VAMP indeed computes an approximation of the posterior mean, by establishing corresponding lower bounds for the mutual information and Bayes-optimal MMSE, together with a TAP characterization of the posterior mean.

We remark that concentration of the overlap between independent samples from the posterior measure may be deduced from the general results of [39] and [40], or alternatively from our calculation of the conditional second moment of a suitably restricted partition function (c.f. Proof of Theorem 1.9 to follow). However, existing arguments that derive the replica-symmetric limit for the free energy from overlap concentration rely on interpolation techniques [9], [41], [42], which seem more difficult to apply in models without an i.i.d. component of the disorder. In this sense, our approach is quite different from the adaptive interpolation method of [14] that is specific to factorized designs $A = BW$ having an i.i.d. Gaussian matrix W. One advantage of our current approach is that it relies less crucially on the Nishimori identities, and is potentially easier to generalize to settings of a mismatched likelihood or prior. Indeed, in sufficiently high-temperature scenarios where a replica-symmetric approximation of the free energy may remain exact even for mismatched models, the results of [25] suggest that a conditional second-moment analysis based on VAMP may still apply, and we leave this as an interesting question to explore in future work.

A. Model and Assumptions

Let $\beta^\star \in \mathbb{R}^n$ be a signal vector with coordinates $(\beta_i^\star)_{i=1}^n \overset{iid}{\sim} \pi$ distributed according to a known prior distribution π. We observe m noisy measurements

$$y = A\beta^\star + \epsilon \in \mathbb{R}^m \tag{1}$$

where $(\epsilon_j)_{j=1}^m \overset{iid}{\sim} N(0,1)$ is Gaussian noise and $A \in \mathbb{R}^{m\times n}$ is the measurement matrix.

Our main results describe the asymptotic mutual information, Bayes-optimal mean squared error, and Bayes posterior-mean estimator for $\beta^\star$, when $n, m \to \infty$ and A is right-rotationally invariant in law. We denote the (normalized) mutual information between $\beta^\star$ and y conditioned on A as

$$i_n := \frac{1}{n} I(\beta^\star; y \mid A) = \frac{1}{n}\mathbb{E}\left(\log \frac{p(y \mid \beta^\star, A)}{p(y \mid A)} \,\middle|\, A\right)$$

We write expectation with respect to the posterior distribution for $\beta^\star$ given (y, A) as $\langle \cdot \rangle$, i.e.

$$\langle f(\sigma)\rangle := \frac{\int f(\sigma)\exp\big(-\frac{1}{2}\|y - A\sigma\|^2\big)\prod_{i=1}^n d\pi(\sigma_i)}{\int \exp\big(-\frac{1}{2}\|y - A\sigma\|^2\big)\prod_{i=1}^n d\pi(\sigma_i)} \tag{2}$$

where we will use σ as the variable for a sample from this posterior. In particular, $\langle\sigma\rangle$ is the posterior mean of β^*. We denote its normalized mean squared error conditioned on A as

$$\mathrm{mmse}_n := \frac{1}{n}\mathbb{E}[\|\beta^\star - \langle\sigma\rangle\|^2 \mid A]. \tag{3}$$

We fix a random variable $\mathsf{D} \geq 0$ representing the limit singular value distribution of A, and denote throughout

$$d_* := \mathbb{E}[\mathsf{D}^2], \qquad d_- := \min(x : x \in \mathrm{supp}(\mathsf{D}^2)),$$
$$d_+ := \max(x : x \in \mathrm{supp}(\mathsf{D}^2))$$

where $\mathrm{supp}(\mathsf{D}^2) \subseteq [0, \infty)$ is the support of D^2.

Assumption 1.1 (Singular Value Distribution): D^2 has strictly positive mean and variance and compact support.

Assumption 1.2 (Measurement Matrix): Let $A = Q^\top DO$ be the singular value decomposition, where $Q \in \mathbb{R}^{m\times m}$ and $O \in \mathbb{R}^{n\times n}$ are orthogonal and $D \in \mathbb{R}^{m\times n}$ is diagonal. Then Q, D are deterministic, $O, \beta^\star, \epsilon$ are mutually independent, and $O \sim \mathrm{Haar}(\mathbb{SO}(n))$ is uniformly distributed on the special orthogonal group. As $n, m \to \infty$,

$$D^\top 1_{m\times 1} \overset{W}{\to} \mathsf{D}, \quad \min\big(\mathrm{diag}(D^\top D)\big) \to d_-,$$
$$\max\big(\mathrm{diag}(D^\top D)\big) \to d_+. \tag{4}$$

Here, $D^\top 1_{m\times 1} \overset{W}{\to} \mathsf{D}$ denotes Wasserstein-$\mathfrak{p}$ convergence of the empirical distribution of coordinates of $D^\top 1_{m\times 1} \in \mathbb{R}^n$

to D for all orders $\mathfrak{p} \geq 1$, and we review properties of this convergence in Appendix G-A.

We may assume without loss of generality that π has mean 0, by subtracting from y a multiple of $A1_{n\times 1}$. Our results are then proven under the following additional assumptions for π and "high-temperature" condition for D.

Assumption 1.3 (Prior distribution): Let $\mathsf{X}^\star \sim \pi$. Then π is a non-Gaussian distribution with

$$\mathbb{E}[\mathsf{X}^\star] = 0, \qquad \rho_* := \mathbb{E}[\mathsf{X}^{\star 2}] > 0.$$

There is a constant $\mathfrak{C} > 0$ for which, for any $s > 0$,

$$\rho_* \leq \mathfrak{C}, \qquad \mathbb{P}[|\mathsf{X}^\star| > s] \leq 2e^{-s^2/(2\mathfrak{C})}. \tag{5}$$

Furthermore, for any $k \in \{1,2\}$, symmetric $\Gamma \in \mathbb{R}^{k\times k}$ satisfying $\Gamma \prec (4\mathfrak{C})^{-1}I$, and $z \in \mathbb{R}^k$, denote

$$d\mu(x) = \frac{e^{x^\top \Gamma x + x^\top z}\prod_{i=1}^k d\pi(x_i)}{\int e^{x^\top \Gamma x + x^\top z}\prod_{i=1}^k d\pi(x_i)},$$
$$\langle f(x)\rangle_\mu = \int f(x)d\mu(x), \quad \mathbb{V}_\mu[f(x)] = \langle f(x)^2\rangle_\mu - \langle f(x)\rangle_\mu^2.$$

Then the distribution μ satisfies, for any unit vector $v \in \mathbb{R}^k$ and a constant $C > 0$ depending only on $\mathfrak{C}$,

$$\mathbb{V}_\mu[v^\top x] \leq C, \qquad \mathbb{V}_\mu[(v^\top x)^2] \leq C[1 + \langle (v^\top x)^2\rangle_\mu]. \tag{6}$$

Assumption 1.4 (High temperature): We have $\mathrm{supp}(\mathsf{D}^2) \subseteq [d_* - \mathfrak{e}, d_* + \mathfrak{e}]$ for some $\mathfrak{e} > 0$.

The condition (6) of Assumption 1.3 may be understood as a Poincaré-type inequality for μ, and holds (for example) when π has a bounded support contained in $[-\sqrt{\mathfrak{C}}, \sqrt{\mathfrak{C}}]$ or a log-concave density $e^{-g(x)}$ where $g''(x) \geq 1/\mathfrak{C}$, c.f. Proposition 13.

In our main results, we will require the value $\mathfrak{e}$ in Assumption 1.4 to be sufficiently small (depending on the constant $\mathfrak{C}$ in Assumption 1.3), and this is the main restriction of our current work. Such a requirement will limit our results to a subset of the regime where the state evolution of VAMP has a unique fixed point, and there is no statistical-computational gap. Note that in a model $y = A\beta^\star + \sigma\epsilon$ with general noise variance $\sigma^2 > 0$, this high temperature assumption encompasses the setting of sufficiently large σ^2 for any fixed singular value distribution D and fixed prior π. (This follows upon rescaling y, A, and D all by $1/\sigma$.)

Remark 1.5: We restrict to non-Gaussian priors π to avoid a rank degeneracy in our subsequent conditioning arguments. If π is Gaussian, our proofs may be modified to condition on only a single iteration of VAMP, rather than t iterations for $t \to \infty$. We will not discuss this modification because i_n, mmse_n, and the posterior mean $\langle\sigma\rangle$ all have explicit formulas for Gaussian priors π, and in this case the main results may be shown at any temperature using more direct techniques of asymptotic random matrix theory (see e.g. [13, Theorem 2] and [43, Theorem 1]).

Remark 1.6: Our results extend directly to the more commonly studied setting where $O \sim \mathrm{Haar}(\mathbb{O}(n))$ is uniform over the full orthogonal group, and also where Q, D are random and independent of $O, \beta^\star, \epsilon$ such that (4) holds almost surely as $n, m \to \infty$. We discuss this further in Appendix G-E.

B. Scalar Channel and Fixed Point Equation

The asymptotic characterization of the model (1) is described by a "single-letter" scalar channel

$$\mathsf{Y} = \mathsf{X}^\star + \mathsf{Z}/\sqrt{\gamma} \tag{7}$$

with signal $\mathsf{X}^\star \sim \pi$, independent Gaussian noise $\mathsf{Z} \sim N(0,1)$, and noise variance $\gamma^{-1} > 0$. We denote the Bayes posterior-mean denoiser in this model as

$$f(y,\gamma) = \mathbb{E}[\mathsf{X}^\star \mid \mathsf{Y} = y] \tag{8}$$

and the signal-observation mutual information and Bayes-optimal mean squared error as

$$i(\gamma) = I(\mathsf{X}^\star;\mathsf{Y}), \qquad \mathrm{mmse}(\gamma) = \mathbb{E}[\mathbb{V}[\mathsf{X}^\star \mid \mathsf{Y}]]. \tag{9}$$

Under Assumption 1.1, let $G : (-d_-, \infty) \to (0,\infty)$ and $R : (0, G(-d_-)) \to (-\infty, 0)$ be the Cauchy- and R-transforms of the law of $-\mathsf{D}^2$, defined by

$$G(z) = \mathbb{E}\left[\frac{1}{z + \mathsf{D}^2}\right], \qquad R(z) = G^{-1}(z) - \frac{1}{z}, \tag{10}$$

where $G^{-1}(\cdot)$ is the functional inverse of $G(\cdot)$, and we set $G(-d_-) = \lim_{z\to -d_-} G(z)$. Lemma 13 shows that these functions are well-defined and reviews several additional properties under the high-temperature condition of Assumption 1.4. The noise variance γ^{-1} that relates this scalar channel to the model (1) is a solution of the fixed-point equations

$$\eta^{-1} = \mathrm{mmse}(\gamma), \qquad \gamma = -R(\eta^{-1}). \tag{11}$$

The second equation can be written equivalently as $\eta^{-1} = G(\eta - \gamma)$ by the definition of the R-transform.

The following ensures that this fixed-point system has a unique solution when $\mathfrak{e}$ in Assumption 1.4 is sufficiently small; see Appendix B-A for its proof.

Proposition 1: Under Assumptions 1.1, 1.3, and 1.4, there exists a constant $\mathfrak{e}_0 = \mathfrak{e}_0(\mathfrak{C}) > 0$ such that if $\mathfrak{e} < \mathfrak{e}_0$, then (11) has a unique solution (η_*^{-1}, γ_*) in the domain $(0, G(-d_-)) \times \mathbb{R}_+$. Furthermore $\eta_*^{-1} \leq \rho_*$ and $\eta_* - \gamma_* \geq \rho_*^{-1} > 0$ where ρ_* is the prior variance in Assumption 1.3.

C. Main Results

Let us denote (η_*^{-1}, γ_*) as the unique fixed point of (11). Our first result describes the asymptotic mutual information in the model (1). We define the replica symmetric potential following [14],

$$i_{\mathrm{RS}}\left(\eta^{-1},\gamma\right) = i(\gamma) - \frac{1}{2}\int_0^{\eta^{-1}} R(z)dz - \frac{\gamma}{2\eta} \tag{12}$$

where $i(\gamma)$ is the above mutual information in the scalar channel.

Theorem 1.7 (Mutual information): Under Assumptions 1.1–1.4, there exists a constant $\mathfrak{e}_0 = \mathfrak{e}_0(\mathfrak{C}) > 0$ such that if $\mathfrak{e} < \mathfrak{e}_0$, then almost surely

$$\lim_{n,m\to\infty} i_n = i_{\mathrm{RS}}\left(\eta_*^{-1}, \gamma_*\right) \tag{13}$$

Remark 1.8: Without the high-temperature condition of Assumption 1.4, the general conjecture [14] is that

$\lim_{n,m\to\infty} i_n = \inf i_{\mathrm{RS}}(\eta^{-1},\gamma)$ where the infimum ranges over all (η^{-1},γ) that solves (11).

Next, we characterize the limiting minimum mean squared error.

Theorem 1.9 (MMSE): Under Assumptions 1.1–1.4, there exists a constant $\mathfrak{e}_0 = \mathfrak{e}_0(\mathfrak{C}) > 0$ such that if $\mathfrak{e} < \mathfrak{e}_0$, then almost surely

$$\lim_{n,m\to\infty} \mathrm{mmse}_n = \eta_*^{-1} \tag{14}$$

Finally, we show that the posterior mean $\langle\sigma\rangle$ for $\beta^\star$ in the model (1) approximately satisfies a system of mean-field equations predicted by the Plefka expansion [24, Eqs. (128-129)]. These equations are an analogue of the Thouless-Anderson-Palmer (TAP) equations for the Sherrington-Kirkpatrick model [44], and of their generalization to orthogonally-invariant spin glass models in [15] and [27].

Theorem 1.10 (TAP equations): Under Assumptions 1.1–1.4, there exists a constant $\mathfrak{e}_0 = \mathfrak{e}_0(\mathfrak{C}) > 0$ such that if $\mathfrak{e} < \mathfrak{e}_0$, then almost surely

$$\lim_{n,m\to\infty} \frac{1}{n}\mathbb{E}\left[\left\|\langle\sigma\rangle - f\left(v,\ \gamma_*\right)\right\|^2 \,\middle|\, A\right] = 0 \tag{15}$$

where $v = \langle\sigma\rangle + \gamma_*^{-1}A^\top(y - A\langle\sigma\rangle)$ and $f(\cdot,\gamma)$ is the posterior-mean denoiser applied entrywise to its first argument.

Remark 1.11: For the following two special subclasses of priors satisfying Assumption 1.3: (i) π has a bounded support contained in $[-\sqrt{\mathfrak{C}},\sqrt{\mathfrak{C}}]$ or (ii) π admits a log-concave density $e^{-g(x)}$ with $g''(x) \geq 1/\mathfrak{C}$, an explicit choice of $\mathfrak{e}_0(\mathfrak{C})$ in Theorems 1.7, 1.9 and 1.10 is $\mathfrak{e}_0(\mathfrak{C}) = \frac{\mathfrak{a}}{\mathfrak{C}}$ for some absolute constant $\mathfrak{a} > 0$. See Section B-B for a justification.

Notation: Denote by $\|\cdot\|$ the ℓ_2-norm for vectors and $\ell_2 \to \ell_2$ operator norm for matrices. For scalars $x_1,\ldots,x_k \in \mathbb{R}$, we write $(x_1,\ldots,x_k) \in \mathbb{R}^k$ to denote a column vector with these entries. For vectors $x_1,\ldots,x_k \in \mathbb{R}^n$, we write $(x_1,\ldots,x_k) \in \mathbb{R}^{n\times k}$ as the matrix containing these columns. $1_{m\times n} \in \mathbb{R}^{m\times n}$ denotes the all-1's matrix, $I_{n\times n}$ denotes the identity matrix, $\succ$ and $\succeq$ denote the positive-definite ordering for matrices, $\mathbb{E}$ and $\mathbb{V}$ denote the expectation and variance of a random variable, and $\mathbb{R}_+ = (0,\infty)$ is the positive real line.

Throughout, we treat $\mathfrak{C}$ in Assumption 1.3 as constant, and we write $x = O(y)$ to mean $|x| \leq Cy$ for a constant $C > 0$ depending only on $\mathfrak{C}$. (In particular, this constant does not depend on d_* or on the small parameter $\mathfrak{e}$ in Assumption 1.4, and we will explicitly track the dependence of various quantities on $d_*, \mathfrak{e}$.)

II. Vector AMP

We first review a version of the VAMP algorithm proposed by [21]. Let $r_2^1 \in \mathbb{R}^n$ be an initialization vector such that there exist random variables $(\mathsf{R}_2^1, \mathsf{X}^\star)$ for which, almost surely as $n,m \to \infty$,

$$(r_2^1, \beta^\star) \stackrel{W}{\to} (\mathsf{R}_2^1, \mathsf{X}^\star), \qquad \gamma_{2,1}^{-1} := \mathbb{E}[(\mathsf{R}_2^1 - \mathsf{X}^\star)^2] > 0. \tag{16}$$

Define from $\gamma_{2,1}$ a sequence of state evolution parameters for $t = 1,2,3,\ldots$

$$\eta_{2,t} = 1/G(\gamma_{2,t}), \quad \gamma_{1,t} = \eta_{2,t} - \gamma_{2,t},$$
$$\eta_{1,t+1} = 1/\mathrm{mmse}(\gamma_{1,t}), \quad \gamma_{2,t+1} = \eta_{1,t+1} - \gamma_{1,t} \tag{17}$$

where $\mathrm{mmse}(\cdot)$ is the scalar channel MMSE from (9) and $G(\cdot)$ is the Cauchy-transform of $-\mathsf{D}^2$ from (10). Now consider the sequence of iterates in $\mathbb{R}^n$, for $t = 1,2,3,\ldots$

$$r_1^t = \frac{1}{\gamma_{1,t}}\left[\eta_{2,t}w^t - \gamma_{2,t}r_2^t\right] \tag{18a}$$
$$r_2^{t+1} = \frac{1}{\gamma_{2,t+1}}\left[\eta_{1,t+1}f\left(r_1^t,\gamma_{1,t}\right) - \gamma_{1,t}r_1^t\right] \tag{18b}$$

where $w^t = \left(A^\top A + \gamma_{2,t}I\right)^{-1}\left(A^\top y + \gamma_{2,t}r_2^t\right)$ and $f(\cdot,\gamma)$ is the posterior-mean denoiser from (8) applied entrywise. This coincides with the VAMP algorithm in [21], specialized to the setting with matched MMSE denoiser, and replacing empirically estimated versions of the parameters $\gamma_{1,t},\eta_{1,t},\gamma_{2,t},\eta_{2,t}$ with their large system limits as defined by (17). The following statement is implied by [21, Theorems 1 and 2]; we check the conditions needed for these results in Appendix A.

Theorem 2.1 [21]: Suppose Assumptions 1.1–1.3 hold, and r_2^1 is independent of (A,ϵ) and satisfies (16). Then each value $\eta_{2,t},\gamma_{1,t},\eta_{1,t+1},\gamma_{2,t+1}$ for $t \geq 1$ is well-defined by (17) and strictly positive. For any 2-pseudo-Lipschitz test function $g : \mathbb{R}^2 \to \mathbb{R}$ and each fixed $t \geq 1$, almost surely

$$\lim_{n,m\to\infty} \frac{1}{n}\sum_{i=1}^n g((r_1^t)_i, \beta_i^\star) = \mathbb{E}[g(\mathsf{R}_1^t, \mathsf{X}^\star)] \tag{19}$$

where $\mathsf{R}_1^t = \mathsf{X}^\star + \mathsf{Z}/\sqrt{\gamma_{1,t}}$ and $\mathsf{Z} \sim N(0,1)$ is independent of $\mathsf{X}^\star$. Furthermore, set

$$\hat\beta_2^t = \left(A^\top A + \gamma_{2,t}I\right)^{-1}\left(A^\top y + \gamma_{2,t}r_2^t\right), \ \hat\beta_1^{t+1} = f(r_1^t,\gamma_{1,t}).$$

Then for each fixed $t \geq 1$, almost surely

$$\lim_{n,m\to\infty} \frac{1}{n}\|\hat\beta_2^t - \beta^\star\|^2 = \eta_{2,t}^{-1},$$
$$\lim_{n,m\to\infty} \frac{1}{n}\|\hat\beta_1^{t+1} - \beta^\star\|^2 = \eta_{1,t+1}^{-1}. \tag{20}$$

Our proofs will use an extended state evolution for a version of this algorithm in a reparametrized form. Letting (η_*^{-1},γ_*) be a fixed point of (11), it is computationally convenient to specialize to a "stationary" initialization of this algorithm given by

$$r_1^0 = \beta^\star + p^0, \qquad \gamma_{1,0} = \gamma_* \tag{21}$$

where $(p_i^0)_{i=1}^n \stackrel{iid}{\sim} N(0,\gamma_*^{-1})$ is independent of all other randomness in the model. (The quantities $r_2^1,\gamma_{2,1}$ in (16) are then defined from this initialization by (17) and (18).) In the following results, we reparametrize the algorithm initialized by (21) and describe its state evolution; proofs are deferred to Appendix A.

Set

$$\Lambda = \frac{\eta_* - \gamma_*}{\gamma_*}\left[\eta_*(D^\top D + (\eta_* - \gamma_*)I)^{-1} - I\right] \in \mathbb{R}^{n\times n},$$
$$\xi = Q\epsilon, \qquad e_b = \frac{\eta_*}{\gamma_*}(D^\top D + (\eta_* - \gamma_*)I)^{-1}D^\top\xi,$$
$$e = O^\top e_b, \tag{22}$$

and define from (18) the new variables in $\mathbb{R}^n$

$$x^t = r_2^t - \beta^\star, \qquad y^t = r_1^t - e - \beta^\star.$$

From the posterior-mean function $f(\cdot,\gamma)$ in (8), define $F : \mathbb{R}^2 \to \mathbb{R}$ as

$$F(p,\beta)=\frac{\eta_*}{\eta_*-\gamma_*}f(p+\beta,\gamma_*)-\frac{\gamma_*}{\eta_*-\gamma_*}p-\frac{\eta_*}{\eta_*-\gamma_*}\beta. \quad (23)$$

Proposition 2: If (η_*^{-1},γ_*) is a fixed point of (11) and $\gamma_{1,0}=\gamma_*$, then $\eta_{1,t}=\eta_{2,t}=\eta_*$, $\gamma_{1,t}=\gamma_*$, and $\gamma_{2,t}=\eta_*-\gamma_*$ for all $t\geq 1$. Furthermore, the VAMP algorithm (18) initialized with (21) is equivalent to the initialization $x^1=F(p^0,\beta^\star)$ and the iterations, for $t=1,2,3,\ldots$

$$s^t=Ox^t, \qquad y^t=O^\top\Lambda s^t, \qquad x^{t+1}=F(y^t+e,\beta^\star). \quad (24)$$

We define the important scalar parameters

$$\begin{aligned}\delta_* &= \frac{1}{\eta_*-\gamma_*}, \quad \sigma_*^2=\delta_*\kappa_*,\\ \kappa_* &= \left(\frac{\eta_*-\gamma_*}{\eta_*}\right)^2\left[\mathbb{E}\frac{\eta_*^2}{(\mathsf{D}^2+\eta_*-\gamma_*)^2}-1\right],\\ b_* &= \frac{1}{\gamma_*}-\frac{\kappa_*}{\eta_*-\gamma_*}. \end{aligned} \quad (25)$$

Let us collect

$$H=\left(\beta^\star, D^\top 1_{m\times 1}, D^\top\xi, \operatorname{diag}(\Lambda), e_b, e, p^0\right)\in\mathbb{R}^{n\times 7}. \quad (26)$$

The following results describe the limit empirical distribution of these quantities constituting H, as well as the state evolution of the iterates x^t,s^t,y^t defined by (24).

Proposition 3: Suppose Assumptions 1.1–1.3 hold. Define random variables

$$\Xi\sim N(0,1), \qquad \mathsf{X}^\star\sim\pi, \qquad \mathsf{P}_0\sim N(0,\gamma_*^{-1}),$$
$$\mathsf{E}\sim N(0,b_*)$$

independent of each other and of D, and set

$$\mathsf{L}=\frac{\eta_*-\gamma_*}{\gamma_*}\left(\frac{\eta_*}{\mathsf{D}^2+\eta_*-\gamma_*}-1\right),$$
$$\mathsf{E}_b=\frac{\eta_*}{\gamma_*}\frac{\mathsf{D}\Xi}{\mathsf{D}^2+\eta_*-\gamma_*},$$
$$\mathsf{H}=(\mathsf{X}^\star,\mathsf{D},\mathsf{D}\Xi,\mathsf{L},\mathsf{E}_b,\mathsf{E},\mathsf{P}_0).$$

Then $\kappa_*=\mathbb{E}\mathsf{L}^2$ and $b_*=\mathbb{E}\mathsf{E}_b^2$. Furthermore, $H\overset{W}{\to}\mathsf{H}$ almost surely as $n,m\to\infty$.

Theorem 2.2: Suppose Assumptions 1.1–1.3 hold. Let $\mathsf{H}=(\mathsf{X}^\star,\mathsf{D},\mathsf{D}\Xi,\mathsf{L},\mathsf{E}_b,\mathsf{E},\mathsf{P}_0)$ be as defined in Proposition 3. Set $\mathsf{X}_1=F(\mathsf{P}_0,\mathsf{X}^\star)$ for the function $F(\cdot)$ from (23), set $\Delta_1=\mathbb{E}[\mathsf{X}_1^2]\in\mathbb{R}^{1\times 1}$, and define iteratively $\mathsf{S}_t,\mathsf{Y}_t,\mathsf{X}_{t+1},\Delta_{t+1}$ for $t=1,2,3,\ldots$ such that

$$(\mathsf{S}_1,\ldots,\mathsf{S}_t)\sim N(0,\Delta_t), \qquad (\mathsf{Y}_1,\ldots,\mathsf{Y}_t)\sim N(0,\kappa_*\Delta_t)$$

are Gaussian vectors independent of each other and of H, and

$$\mathsf{X}_{t+1}=F(\mathsf{Y}_t+\mathsf{E},\mathsf{X}^\star),$$
$$\Delta_{t+1}=\mathbb{E}\left[(\mathsf{X}_1,\ldots,\mathsf{X}_{t+1})(\mathsf{X}_1,\ldots,\mathsf{X}_{t+1})^\top\right]\in\mathbb{R}^{(t+1)\times(t+1)}.$$

Then for each $t\geq 1$, $\Delta_t\succ 0$ strictly, $\delta_*=\mathbb{E}\mathsf{X}_t^2$, and $\sigma_*^2=\mathbb{E}\mathsf{Y}_t^2$.

Furthermore, let $X_t=(x^1,\ldots,x^t)\in\mathbb{R}^{n\times t}$, $S_t=(s^1,\ldots,s^t)\in\mathbb{R}^{n\times t}$, and $Y_t=(y^1,\ldots,y^t)\in\mathbb{R}^{n\times t}$ collect the iterates of (24), starting from the initialization $x^1=F(p^0,\beta^\star)$. Then for any fixed $t\geq 1$, almost surely as $n,m\to\infty$,

$$(H,X_t,S_t,Y_t)\overset{W}{\to}(\mathsf{H},\mathsf{X}_1,\ldots,\mathsf{X}_t,\mathsf{S}_1,\ldots,\mathsf{S}_t,\mathsf{Y}_1,\ldots,\mathsf{Y}_t).$$

Corollary 1: In the setting of Theorem 2.2, for any fixed $t\geq 1$, almost surely

$$\lim_{n,m\to\infty} n^{-1}(e,X_t,Y_t)^\top(e,X_t,Y_t)=\begin{pmatrix} b_* & 0 & 0\\ 0 & \Delta_t & 0\\ 0 & 0 & \kappa_*\Delta_t\end{pmatrix}$$

Theorem 2.2 implies that the joint limit $(\mathsf{S}_1,\ldots,\mathsf{S}_t)$ for the iterates $S_t=(s^1,\ldots,s^t)$ is independent of D. We highlight here the following implication, which is an analogue of [33, Proposition 2.4].

Corollary 2: In the setting of Theorem 2.2, fix any $t\geq 1$, let $f:\mathbb{R}\to\mathbb{R}$ be any function which is continuous and bounded in a neighborhood of $\operatorname{supp}(\mathsf{D}^2)$, and define $f(D^\top D)\in\mathbb{R}^{n\times n}$ by the functional calculus. Then almost surely

$$\lim_{n,m\to\infty} n^{-1}S_t^\top f(D^\top D)S_t=\Delta_t\cdot\mathbb{E}f(\mathsf{D}^2).$$

Noting that each matrix Δ_t is the upper-left submatrix of Δ_{t+1}, let us denote the entries of these matrices as $\Delta_t=(\delta_{rs})_{r,s=1}^t$. Theorem 2.2 ensures that $\delta_{tt}=\delta_*$ for all $t\geq 1$. The following result then guarantees that for sufficiently small $\mathfrak{e}$ in Assumption 1.4, the state evolution of this stationary VAMP algorithm is convergent in the sense

$$\lim_{\min(s,t)\to\infty}\left(\lim_{n,m\to\infty}\frac{1}{n}\left\|x^t-x^s\right\|^2\right)=\lim_{\min(s,t)\to\infty}(\delta_{ss}+\delta_{tt}-2\delta_{st})=0$$
$$\lim_{\min(s,t)\to\infty}\left(\lim_{n,m\to\infty}\frac{1}{n}\left\|y^t-y^s\right\|^2\right)=\lim_{\min(s,t)\to\infty}\kappa_*(\delta_{ss}+\delta_{tt}-2\delta_{st})=0.$$

Proposition 4: Under Assumptions 1.1–1.4, there exists some constant $\mathfrak{e}_0=\mathfrak{e}_0(\mathfrak{C})>0$ such that if $\mathfrak{e}<\mathfrak{e}_0$, then $\lim_{\min(s,t)\to\infty}\delta_{st}=\delta_*$.

We verify in Appendix B-D that, outside the high-temperature regime of Assumption 1.4, this statement of Proposition 4 continues to hold for the stationary initialization of VAMP defined by any fixed point (η_*^{-1},γ_*) to (11) that is a local minimizer of the replica-symmetric potential (12).

III. Analysis of the Restricted Partition Function

For any subset $\mathcal{U}\subseteq[0,\infty)$, define a restricted partition function for the model (1) as

$$\mathcal{Z}(\mathcal{U})=\int\mathbb{I}\left(\frac{1}{n}\|\sigma-\beta^\star\|^2\in\mathcal{U}\right)\cdot\exp\left(-\frac{\|y-A\sigma\|^2}{2}\right)\prod_{i=1}^n d\pi(\sigma_i). \quad (27)$$

We will ultimately analyze the unrestricted partition function $\mathcal{Z}=\mathcal{Z}([0,\infty))$, although it is technically convenient to first analyze $\mathcal{Z}(\mathcal{U})$ for bounded subsets $\mathcal{U}$.

For any $t \geq 1$, define the sigma-field (in the probability space of O, $\beta^\star$, and ϵ)

$$\mathcal{G}_t = \mathcal{G}\left(H, x^1, s^1, y^1, \ldots, x^t, s^t, y^t\right) \quad (28)$$

where H consists of the quantities in (26), and x^t, s^t, y^t are the VAMP iterates of (24). In this section, we provide asymptotic variational characterizations of the conditional first and second moments $\mathbb{E}[\mathcal{Z}(\mathcal{U}) \mid \mathcal{G}_t]$ and $\mathbb{E}[\mathcal{Z}(\mathcal{U})^2 \mid \mathcal{G}_t]$. Together with a concentration inequality for $\log \mathcal{Z}(\mathcal{U})$ and a second-moment argument, these establish an unconditional first-order limit for $\log \mathcal{Z}(\mathcal{U})$. Proofs of these results are provided in Appendices C, D, and E.

For $a, b \in \mathbb{R}$ and $M > 0$, define

$$\begin{aligned} c_\pi(a,b) &= \int \exp\left(ax^2 + bx\right) d\pi(x) \in (0, \infty], \\ c_\pi^M(a,b) &= \int_{-M}^{M} \exp\left(ax^2 + bx\right) d\pi(x) \in (0, \infty). \end{aligned} \quad (29)$$

When π has unbounded support, $c_\pi(a,b)$ may be infinite for large positive values of a, and we discuss its behavior in Lemma 9. Under Assumption 1.4, recall the unique fixed point (η_*^{-1}, γ_*) of (11) and the prior variance ρ_* from Assumption 1.3, and define the replica-symmetric free energy

$$\begin{aligned} \Psi_{\mathrm{RS}} &= -\frac{1}{2} - \frac{\gamma_*\rho_*}{2} + \frac{\gamma_*}{2\eta_*} + \frac{1}{2}\int_0^{\eta_*^{-1}} R(z)dz \\ &\quad + \mathbb{E}\log c_\pi\left(-\frac{\gamma_*}{2}, \gamma_* \mathsf{X}^\star + \sqrt{\gamma_*}\mathsf{Z}\right) \end{aligned} \quad (30)$$

where the expectation is over independent variables $\mathsf{X}^\star \sim \pi$ and $\mathsf{Z} \sim N(0,1)$. In addition to d_*, ρ_* and the variance parameters $\delta_*, \kappa_*, \sigma_*^2, b_*$ from (25), we introduce the auxiliary scalar parameters

$$\begin{aligned} a_* &= (\eta_* - \gamma_*)\left(1 - \frac{d_*}{\gamma_*}\right), \\ c_* &= -(\eta_* - \gamma_*)\kappa_* + \left(\frac{\eta_* - \gamma_*}{\gamma_*}\right)^2 (d_* - \gamma_*), \\ e_* &= 1 + \kappa_*, \quad \alpha_*^A = \frac{\eta_* - \gamma_*}{\gamma_*}\frac{1}{\sqrt{\kappa_*}}, \\ \alpha_*^B &= \left(\alpha_*^A\right)^2 (\gamma_* - d_*), \quad \pi_* = \mathbb{E}\mathsf{X}^\star F\left(\frac{\mathsf{Z}}{\sqrt{\gamma_*}}, \mathsf{X}^\star\right). \end{aligned} \quad (31)$$

A. Conditional First Moment

Fix an iteration $t \geq 1$ for VAMP, and define tuples of primal and dual variables

$$\mathfrak{P} = (u, r, v, w), \qquad \mathfrak{Q} = (\zeta, U, R, V, W, \chi^A, \chi^B, \chi^C)$$

where $u > 0$ and $\zeta > -d_-$ and $r, U, R, \chi^A, \chi^B, \chi^C \in \mathbb{R}$ and $v, w, V, W \in \mathbb{R}^t$. Define

$$\mathcal{A}(\mathfrak{P}) = u - r^2 - \|v\|^2 - \|w\|^2, \quad \mathcal{V} = \{\mathfrak{P} : \mathcal{A}(\mathfrak{P}) > 0\}. \quad (32)$$

Define also the functions

$$\begin{aligned} \mathcal{H}(\zeta, \mathcal{A}) &= \zeta\mathcal{A} - \mathbb{E}[\log(\zeta + \mathsf{D}^2)] - (1 + \log \mathcal{A}), \\ \mathcal{B}(v,w) &= \|v - \alpha_*^A w\|^2 \\ \lambda(x) &= \frac{\eta_* - \gamma_*}{\gamma_*}\left(\frac{\eta_*}{x + \eta_* - \gamma_*} - 1\right), \\ \theta(x) &= x - \frac{\alpha_*^B \eta_*}{x + \eta_* - \gamma_*} + \alpha_*^B - d_*. \end{aligned} \quad (33)$$

Let Δ_t and the random variables $(\mathsf{E}, \mathsf{X}^\star, \mathsf{X}_1, \ldots, \mathsf{X}_t, \mathsf{Y}_1, \ldots, \mathsf{Y}_t)$ be as described in Proposition 3 and Theorem 2.2, and define

$$\begin{aligned} \Phi_{1,t}(\mathfrak{P}, \mathfrak{Q}) &= -\frac{1}{2} + \mathbb{E}\log c_\pi\Bigg(U, -2U\mathsf{X}^\star + \frac{R\,\mathsf{E}}{\sqrt{b_*}} \\ &\quad + V^\top \Delta_t^{-1/2}(\mathsf{X}_1, \ldots, \mathsf{X}_t) + \frac{W^\top \Delta_t^{-1/2}(\mathsf{Y}_1, \ldots, \mathsf{Y}_t)}{\sqrt{\kappa_*}}\Bigg) \\ &\quad - (u - \rho_*)U - r\,R - (v + \pi_*\Delta_t^{-1/2}1_{t\times 1})^\top V - w^\top W \\ &\quad - \frac{1}{2}\Bigg(\frac{e_* r^2}{b_*} - \frac{2r}{\sqrt{b_*}} + \mathrm{Tr}\Bigg[\begin{pmatrix} d_* & a_* \\ a_* & c_* \end{pmatrix}\left(v, \frac{w}{\sqrt{\kappa_*}}\right)^\top \\ &\quad \times \left(v, \frac{w}{\sqrt{\kappa_*}}\right)\Bigg]\Bigg) + \frac{1}{2}\mathcal{H}(\zeta, \mathcal{A}(\mathfrak{P})) \\ &\quad + \frac{1}{2}\mathbb{E}\left[\frac{\mathsf{E}_b^2}{\zeta + \mathsf{D}^2}\left(\left[\frac{\gamma_*}{\eta_*} - \frac{r}{\sqrt{b_*}}\right]\mathsf{D}^2 - \chi^A\right)^2\right] \\ &\quad + \frac{1}{2}\mathbb{E}\left[\frac{1}{\zeta + \mathsf{D}^2}\left(\theta(\mathsf{D}^2) - \lambda(\mathsf{D}^2)\chi^B - \chi^C\right)^2\right]\mathcal{B}(v,w) \end{aligned} \quad (34)$$

Let $\Phi_{1,t}^M(\mathfrak{P}, \mathfrak{Q})$ have the same definition with c_π replaced by c_π^M. Finally, define

$$\begin{aligned} \Psi_{1,t}(\mathfrak{P}) &= \inf_{\mathfrak{Q}:\zeta > -d_-} \Phi_{1,t}(\mathfrak{P}, \mathfrak{Q}), \\ \Psi_{1,t}^M(\mathfrak{P}) &= \inf_{\mathfrak{Q}:\zeta > -d_-} \Phi_{1,t}^M(\mathfrak{P}, \mathfrak{Q}) \end{aligned} \quad (35)$$

where these may take extended real values in $[-\infty, \infty)$.

Lemma 1: Fix any $K > 0$. Under Assumptions 1.1–1.4, there exists a constant $\mathfrak{e}_0 = \mathfrak{e}_0(\mathfrak{C}, K) > 0$ such that if $\mathfrak{e} < \mathfrak{e}_0$, then for any fixed $t \geq 1$ and non-empty open subset $\mathcal{U} \subseteq (0, K)$, almost surely

$$\begin{aligned} \liminf_{n,m\to\infty} \frac{1}{n}\log \mathbb{E}\left[\mathcal{Z}(\mathcal{U}) \mid \mathcal{G}_t\right] &\geq \sup_{\mathfrak{P}\in\mathcal{V}:\, u\in\mathcal{U}} \sup_{M>0} \Psi_{1,t}^M(\mathfrak{P}), \\ \limsup_{n,m\to\infty} \frac{1}{n}\log \mathbb{E}\left[\mathcal{Z}(\overline{\mathcal{U}}) \mid \mathcal{G}_t\right] &\leq \sup_{\mathfrak{P}\in\mathcal{V}:\, u\in\mathcal{U}} \Psi_{1,t}(\mathfrak{P}) \end{aligned}$$

where $\overline{\mathcal{U}}$ is the closure of $\mathcal{U}$.

Lemma 2: Fix any $K > 0$. Under Assumptions 1.1–1.4, there exists a constant $\mathfrak{e}_0 = \mathfrak{e}_0(\mathfrak{C}, K) > 0$ such that if $\mathfrak{e} < \mathfrak{e}_0$ and $\mathcal{U} \subseteq (0, K)$ is any fixed open subset containing $2\eta_*^{-1}$, then

$$\begin{aligned} \liminf_{t\to\infty} \sup_{\mathfrak{P}\in\mathcal{V}:\, u\in\mathcal{U}} \sup_{M>0} \Psi_{1,t}^M(\mathfrak{P}) &\geq \Psi_{\mathrm{RS}}, \\ \limsup_{t\to\infty} \sup_{\mathfrak{P}\in\mathcal{V}:\, u\in\mathcal{U}} \Psi_{1,t}(\mathfrak{P}) &\leq \Psi_{\mathrm{RS}}. \end{aligned} \quad (36)$$

Furthermore, there exists a universal constant $c_0 > 0$ such that for any $\varsigma > 0$,

$$\limsup_{t\to\infty} \sup_{\substack{\mathfrak{P}\in\mathcal{V}:\, u\in\mathcal{U} \\ |u - 2\eta_*^{-1}| > \varsigma}} \Psi_{1,t}(\mathfrak{P}) < \Psi_{\mathrm{RS}} - c_0\mathfrak{e}^{1/2}\varsigma^2. \quad (37)$$

B. Conditional Second Moment

Again fixing an iteration $t \geq 1$ for VAMP, define tuples of primal and dual variables

$$\mathfrak{P} = (u, r, v, w, p), \qquad \mathfrak{Q} = (\mathfrak{Z}, U, R, V, W, P, \chi^A, \chi^B, \chi^C)$$

where $u \in \mathbb{R}_+^2$, $r, U, R \in \mathbb{R}^2$, $v, w, V, W \in \mathbb{R}^{t\times 2}$, $p, P \in \mathbb{R}$, $\mathfrak{Z} \in \mathbb{R}^{2\times 2}$ is symmetric, and $\chi^A, \chi^B, \chi^C \in \mathbb{R}^2$. We write $v = (v_1, v_2)$ where $v_1, v_2 \in \mathbb{R}^t$ are its columns, and similarly for w, V, W. Set the domain for $\mathfrak{Z}$ as

$$\mathcal{D}_+ = \{\mathfrak{Z} \in \mathbb{R}^{2\times 2} : \mathfrak{Z} = \mathfrak{Z}^\top,\ \mathfrak{Z} \succ -d_- \cdot I_{2\times 2}\}, \tag{38}$$

and denote the eigen-decomposition of $\mathfrak{Z}$ by

$$\mathfrak{Z} = \begin{pmatrix} y_1 & y_2 \end{pmatrix} \begin{pmatrix} \zeta_1 & 0 \\ 0 & \zeta_2 \end{pmatrix} \begin{pmatrix} y_1^\top \\ y_2^\top \end{pmatrix} \tag{39}$$

where $y_1, y_2 \in \mathbb{R}^2$ are unit-norm eigenvectors of $\mathfrak{Z}$ and ζ_1, ζ_2 are the corresponding eigenvalues. (The expressions below do not depend on the signs of y_1, y_2 or the choice of y_1, y_2 when $\zeta_1 = \zeta_2$.) Set

$$\mathcal{A}(\mathfrak{P}) = \begin{pmatrix} u_1 & p \\ p & u_2 \end{pmatrix} - rr^\top - v^\top v - w^\top w \in \mathbb{R}^{2\times 2},$$
$$\mathcal{V} = \{\mathfrak{P} : \mathcal{A}(\mathfrak{P}) \succ 0\}. \tag{40}$$

Let $\theta(x), \lambda(x)$ be as in (33), and define for $a, b \in \mathbb{R}^2$ and $c \in \mathbb{R}$

$$c_\pi(a, b, c) = \int \exp\big(a_1 x_1^2 + b_1 x_1 \tag{41}$$
$$+ a_2 x_2^2 + b_2 x_2 + c x_1 x_2\big) d\pi(x_1) d\pi(x_2) \tag{42}$$

$$\mathcal{H}(\mathfrak{Z}, \mathcal{A}) = \mathrm{Tr}(\mathfrak{Z}\mathcal{A}) - \mathbb{E}\Big[\log\det\big(\mathfrak{Z} + \mathsf{D}^2 \cdot I_{2\times 2}\big)\Big] - (2 + \log\det \mathcal{A})$$
$$\mathcal{B}(v, w) = (v - \alpha_*^A w)^\top (v - \alpha_*^A w) \tag{43}$$

Here $c_\pi(p, a, b), \mathcal{H}(\mathfrak{Z}, \mathcal{A}) \in \mathbb{R}$ and $\mathcal{B}(v, w) \in \mathbb{R}^{2\times 2}$. Finally, define from the above (44), shown at the bottom of the next page, and denote

$$\Psi_{2,t}(\mathfrak{P}) = \inf_{\mathfrak{Q}: \mathfrak{Z} \in \mathcal{D}_+} \Phi_{2,t}(\mathfrak{P}, \mathfrak{Q}). \tag{45}$$

The following results are analogous to the upper bounds of Lemmas 1 and 2. (Lower bounds may also be shown, but we omit these statements as we require only the upper bounds in the subsequent proofs.)

Lemma 3: Fix any $K > 0$. Under Assumptions 1.1–1.4, there exists a constant $\mathfrak{e}_0 = \mathfrak{e}_0(\mathfrak{C}, K) > 0$ such that if $\mathfrak{e} < \mathfrak{e}_0$, then for any fixed $t \geq 1$ and non-empty open subset $\mathcal{U} \subseteq (0, K)$, almost surely

$$\lim_{n,m\to\infty} \frac{1}{n} \log \mathbb{E}\left[\mathcal{Z}(\overline{\mathcal{U}})^2 \mid \mathcal{G}_t\right] \leq \sup_{\mathfrak{P}\in\mathcal{V}:\, u_1, u_2 \in \mathcal{U}} \Psi_{2,t}(\mathfrak{P}) \tag{46}$$

where $\overline{\mathcal{U}}$ is the closure of $\mathcal{U}$.

Lemma 4: Fix any $K > 0$. Under Assumptions 1.1–1.4, there exists a constant $\mathfrak{e}_0 = \mathfrak{e}_0(\mathfrak{C}, K) > 0$ such that if $\mathfrak{e} < \mathfrak{e}_0$ and $\mathcal{U} \subseteq (0, K)$ is any fixed open set containing $2\eta_*^{-1}$, then

$$\limsup_{t\to\infty} \sup_{\mathfrak{P}\in\mathcal{V}:\, u_1, u_2 \in \mathcal{U}} \Psi_{2,t}(\mathfrak{P}) \leq 2\Psi_{\mathrm{RS}}.$$

C. Limiting Free Energy

Combining the preceding results with the following exponential concentration inequality for $n^{-1}\log \mathcal{Z}(\mathcal{U})$, we deduce as a corollary an unconditional first-order limit for the restricted free energy.

Lemma 5: Fix any $K, L > 0$ and subset $\mathcal{U} \subseteq [0, K]$. Let $\mathcal{E}$ denote the event where

$$\int \mathbb{I}\left(\frac{1}{n}\|\sigma - \beta^\star\|^2 \in \mathcal{U}\right) \prod_{i=1}^n d\pi(\sigma_i) > 0, \quad \|D^\top \xi\|^2 \leq Ln.$$

(Note that $\mathcal{E}$ depends on the random quantities $(O, \beta^\star, \epsilon)$ only via $(\beta^\star, D^\top \xi)$ and is hence $\mathcal{G}_t$-measurable for any $t \geq 1$.) Under Assumptions 1.1 and 1.2, there exists a constant $C(K, L, d_+) > 0$ such that for any $t \geq 1$, any $\delta > 0$, and all sufficiently large n,

$$\mathbb{P}\left(\left|\frac{1}{n}\log \mathcal{Z}(\mathcal{U}) - \mathbb{E}\left[\frac{1}{n}\log \mathcal{Z}(\mathcal{U}) \,\middle|\, \mathcal{G}_t\right]\right| \geq \delta \,\middle|\, \mathcal{G}_t\right) \mathbb{I}\{\mathcal{E}\}$$
$$\leq 2\exp\left(\frac{-\delta^2 n}{C(K, L, d_+)}\right). \tag{47}$$

Corollary 3: Fix any $K > 0$. Under Assumptions 1.1–1.4, there exists a constant $\mathfrak{e}_0 = \mathfrak{e}_0(\mathfrak{C}, K) > 0$ such that if $\mathfrak{e} < \mathfrak{e}_0$ and $2\eta_*^{-1} \in (0, K)$, then almost surely

$$\lim_{n,m\to\infty} \frac{1}{n} \log \mathcal{Z}([0, K]) = \Psi_{\mathrm{RS}}.$$

IV. Proofs of the Main Results

We use the preceding lemmas to prove Theorems 1.7, 1.9, and 1.10. For expositional clarity, we consider in this section the simpler setting where π has compact support contained in $[-\sqrt{\mathfrak{C}}, \sqrt{\mathfrak{C}}]$. We extend these proofs to the more general condition of Assumption 1.3 in Appendix F.

A. Mutual Information

Proof: [Proof of Theorem 1.7, bounded support] Letting $p(y \mid \beta^\star, A)$ and $p(y \mid A)$ be the conditional density functions of $y \in \mathbb{R}^m$, direct calculation yields

$$\mathbb{E}\left[\log\left((2\pi)^{n/2} p(y \mid \beta^\star, A)\right) \,\middle|\, A\right] = \mathbb{E}\left[-\frac{\|y - A\beta^\star\|^2}{2} \,\middle|\, A\right] = -\frac{n}{2},$$
$$\log\left((2\pi)^{n/2} p(y \mid A)\right) = \log \mathcal{Z}$$

where $\mathcal{Z} = \mathcal{Z}([0, \infty))$ is the unrestricted partition function defined by (27). So the normalized mutual information in the model (1) is

$$i_n = \frac{1}{n} I(\beta^\star; y \mid A) = \frac{1}{n}\mathbb{E}\left(\log \frac{p(y \mid \beta^\star, A)}{p(y \mid A)} \,\middle|\, A\right)$$
$$= -\frac{1}{n}\mathbb{E}[\log \mathcal{Z} \mid A] - \frac{1}{2}.$$

Similarly, the mutual information $i(\gamma_*) = I(\mathsf{X}^\star; \mathsf{Y})$ in the scalar channel (7) is

$$i(\gamma_*) = \mathbb{E}\left(\log \frac{p(\mathsf{Y} \mid \mathsf{X}^\star)}{p(\mathsf{Y})}\right)$$
$$= -\frac{1}{2} - \mathbb{E}\log \int \exp\left(-\frac{(\sqrt{\gamma_*}(\mathsf{X}^\star - x) + \mathsf{Z})^2}{2}\right) d\pi(x)$$
$$= -\frac{1}{2} + \mathbb{E}\frac{(Z + \sqrt{\gamma_*}\mathsf{X}^\star)^2}{2} \tag{48}$$
$$-\mathbb{E}\log \int \exp\left(-\frac{1}{2}\gamma_* x^2 + (\gamma_* \mathsf{X}^\star + \sqrt{\gamma_*} Z)\, x\right) d\pi(x)$$
$$= \frac{1}{2}\gamma_*\rho_* - \mathbb{E}\log c_\pi\left(-\frac{1}{2}\gamma_*, \gamma_* \mathsf{X}^\star + \sqrt{\gamma_*} Z\right) \tag{49}$$

where c_π is defined in (29).

Suppose π has bounded support contained in $[-\sqrt{\mathfrak{C}}, \sqrt{\mathfrak{C}}]$. Setting $K = 4\mathfrak{C}$ and $\mathcal{U} = (0, K)$, we have $n^{-1}\|\sigma - \beta^\star\|^2 \le K$ with probability 1. Then $\mathcal{Z}(\overline{\mathcal{U}}) = \mathcal{Z}$, and also $2\eta_*^{-1} \le 2\rho_* < K$ where the first inequality is by Proposition 1. Thus Corollary 3 shows $n^{-1}\log \mathcal{Z} \to \Psi_{\mathrm{RS}}$ almost surely. By Jensen's inequality,

$$0 \le -\frac{1}{n}\log \mathcal{Z} \le \frac{1}{n}\int \frac{\|A(\beta^\star - \sigma) + \epsilon\|^2}{2} \prod_{i=1}^{n} d\pi(\sigma_i)$$
$$\le \left\|A^\top A\right\|_{\mathrm{op}} \cdot \int \frac{\|\beta^\star - \sigma\|^2}{n} \prod_{i=1}^{n} d\pi(\sigma_i) + \frac{1}{n}\|\epsilon\|^2. \tag{50}$$

Then by the given assumption $\|A^\top A\|_{\mathrm{op}} = \|D^\top D\|_{\mathrm{op}} \to d_+$, the bound $n^{-1}\|\beta^\star - \sigma\|^2 \le K$, uniform integrability of $\{\|\epsilon\|^2/n\}_{n \ge 1}$, and the dominated convergence theorem, almost surely $n^{-1}\mathbb{E}[\log \mathcal{Z} \mid A] \to \Psi_{\mathrm{RS}}$. Applying this, (48), and the forms of Ψ_{RS} and i_{RS} in (30) and (12), we obtain as desired

$$\lim_{n,m\to\infty} i_n = -\Psi_{\mathrm{RS}} - 1/2 = i_{\mathrm{RS}}(\eta_*^{-1}, \gamma_*).$$

■

B. Bayes Risk

Lemma 6: Denoting by $\langle f(\sigma)\rangle$ the posterior expectation in (2),

$$\mathrm{mmse}_n = \frac{1}{2n}\mathbb{E}\left[\left\langle \|\sigma - \beta^\star\|^2 \right\rangle \,\middle|\, A\right].$$

Proof: Let σ, τ denote two replicas sampled independently from the posterior distribution defining (2). Conditional on y and A, since $\sigma, \tau, \beta^\star$ are independent and equal in law, we have the Nishimori identity for any integrable function f (see also [14, Appendix A])

$$\mathbb{E}[\langle f(\sigma,\tau)\rangle \mid A] = \mathbb{E}[\mathbb{E}[f(\sigma,\tau) \mid y, A] \mid A]$$
$$= \mathbb{E}[\mathbb{E}[f(\sigma,\beta^\star) \mid y, A] \mid A]$$
$$= \mathbb{E}[\langle f(\sigma,\beta^\star)\rangle \mid A].$$

Thus

$$\begin{aligned} n \cdot \mathrm{mmse}_n &= \mathbb{E}[\|\beta^\star - \langle\sigma\rangle\|^2 \mid A] \\ &= \mathbb{E}[\beta^{\star\top}\beta^\star + \langle\sigma^\top\tau\rangle - \langle\sigma^\top\beta^\star\rangle - \langle\tau^\top\beta^\star\rangle \mid A] \\ &\overset{(a)}{=} \mathbb{E}[\beta^{\star\top}\beta^\star - \langle\sigma^\top\beta^\star\rangle \mid A] \\ &= \mathbb{E}[\langle(\sigma - \beta^\star)^\top(-\beta^\star)\rangle \mid A] \\ &\overset{(b)}{=} \mathbb{E}[\langle(\sigma - \beta^\star)^\top\sigma\rangle \mid A] \end{aligned} \tag{51}$$

$$\begin{aligned}
\Phi_{2,t}(\mathfrak{P},\mathfrak{Q}) = &-1 + \mathbb{E}\log c_\pi\Bigg(U,\ -2U\,\mathsf{X}^\star + \frac{R\,\mathsf{E}}{\sqrt{b_*}} \\
&+ V^\top\Delta_t^{-1/2}(\mathsf{X}_1,\ldots,\mathsf{X}_t) + \frac{W^\top\Delta_t^{-1/2}(\mathsf{Y}_1,\ldots,\mathsf{Y}_t)}{\sqrt{\kappa_*}} - P\,\mathsf{X}^\star 1_{2\times1},\ P\Bigg) \\
&- (u - \rho_* 1_{2\times1})^\top U - r^\top R - (v_1 + \pi_*\Delta_t^{-1/2}1_{t\times1})^\top V_1 5 \\
&- (v_2 + \pi_*\Delta_t^{-1/2}1_{t\times1})^\top V_2 - w_1^\top W_1 - w_2^\top W_2 - (p - \rho_*)P \\
&- \frac{1}{2}\Bigg(\frac{e_*\|r\|^2}{b_*} - \frac{2r^\top 1_{2\times1}}{\sqrt{b_*}} + \mathrm{Tr}\begin{pmatrix} d_* & a_* \\ a_* & c_* \end{pmatrix} \\
&\qquad \times \left[\left(v_1, \frac{w_1}{\sqrt{\kappa_*}}\right)^\top \left(v_1, \frac{w_1}{\sqrt{\kappa_*}}\right) + \left(v_2, \frac{w_2}{\sqrt{\kappa_*}}\right)^\top \left(v_2, \frac{w_2}{\sqrt{\kappa_*}}\right)\right]\Bigg) \\
&+ \frac{1}{2}\mathcal{H}(\mathfrak{Z}, \mathcal{A}(\mathfrak{P})) \\
&+ \frac{1}{2}\mathbb{E}\Bigg[\left(\left[\frac{\gamma_*}{\eta_*}1_{2\times1} - \frac{r}{\sqrt{b_*}}\right]\mathsf{D}^2 - \chi^A\right)^\top \left(\mathsf{E}_b^2(\mathfrak{Z} + \mathsf{D}^2\cdot I_{2\times2})^{-1}\right) \\
&\qquad \times \left(\left[\frac{\gamma_*}{\eta_*}1_{2\times1} - \frac{r}{\sqrt{b_*}}\right]\mathsf{D}^2 - \chi^A\right)\Bigg] \\
&+ \frac{1}{2}\mathbb{E}\left[\frac{\left(\theta(\mathsf{D}^2) - \lambda(\mathsf{D}^2)\chi_1^B - \chi_1^C\right)^2}{\zeta_1 + \mathsf{D}^2}\right] y_1^\top \mathcal{B}(v,w) y_1 \\
&+ \frac{1}{2}\mathbb{E}\left[\frac{\left(\theta(\mathsf{D}^2) - \lambda(\mathsf{D}^2)\chi_2^B - \chi_2^C\right)^2}{\zeta_2 + \mathsf{D}^2}\right] y_2^\top \mathcal{B}(v,w) y_2
\end{aligned} \tag{44}$$

where (a) applies the Nishimori identity, and (b) follows from the exchangeability of the replicas and the Nishimori identity as follows:

$$\begin{aligned}\mathbb{E}[\langle(\sigma-\beta^\star)^\top(-\beta^\star)\rangle \mid A] &= \mathbb{E}[\langle(\sigma-\tau)^\top(-\tau)\rangle \mid A] \\ &= \mathbb{E}[\langle(\tau-\sigma)^\top(-\sigma)\rangle \mid A] \\ &= \mathbb{E}[\langle(\beta^\star-\sigma)^\top(-\sigma)\rangle \mid A].\end{aligned}$$

Summing the last two expressions in (51) gives $2n\cdot\mathrm{mmse}_n = \mathbb{E}[\langle\|\sigma-\beta^\star\|^2\rangle \mid A]$ as required. ∎

Proof: [Proof of Theorem 1.9, bounded support] Suppose again that π has support contained in $[-\sqrt{\mathfrak{C}},\sqrt{\mathfrak{C}}]$, and set $K = 4\mathfrak{C}$. Then again $n^{-1}\|\sigma-\beta^\star\|^2 \le K$ with probability 1, the unrestricted partition function is $\mathcal{Z} = \mathcal{Z}([0,K])$, and $2\eta_*^{-1} \le 2\rho_* < K$. Fix any small constant $\varsigma > 0$ and set $\mathcal{U} = (0,K) \setminus (2\eta_*^{-1}-\varsigma, 2\eta_*^{-1}+\varsigma)$. Then

$$\langle\mathbb{I}\left(\left|\frac{1}{n}\|\sigma-\beta^\star\|^2 - 2\eta_*^{-1}\right| > \varsigma\right)\rangle = \frac{\mathcal{Z}(\overline{\mathcal{U}})}{\mathcal{Z}}.$$

Applying Lemmas 1, 2, and Jensen's inequality, for a sufficiently large iteration $t \ge 1$, almost surely for all large n,

$$\frac{1}{n}\mathbb{E}\left[\log\mathcal{Z}(\overline{\mathcal{U}}) \mid \mathcal{G}_t\right] \le \frac{1}{n}\log\mathbb{E}\left[\mathcal{Z}(\overline{\mathcal{U}}) \mid \mathcal{G}_t\right] < \Psi_{\mathrm{RS}} - c_0\mathfrak{e}^{1/2}\varsigma^2.$$

Taking the expectation of (47) from Lemma 5 yields the unconditional tail bound

$$\mathbb{P}\left(\mathcal{E}\text{ holds and }\left|\frac{1}{n}\log\mathcal{Z}(\overline{\mathcal{U}}) - \mathbb{E}\left[\frac{1}{n}\log\mathcal{Z}(\overline{\mathcal{U}}) \,\middle|\, \mathcal{G}_t\right]\right| \ge \delta\right) \le 2\exp\left(\frac{-\delta^2 n}{C(K,L,d_+)}\right). \tag{52}$$

Applying $\|D^\top D\|_{\mathrm{op}} \to d_+$ as $n\to\infty$ and a standard chi-squared tail bound, for a sufficiently large constant $L > 0$, the second condition $\|D^\top\xi\|^2 \le Ln$ defining $\mathcal{E}$ holds almost surely for all large n. The first condition defining $\mathcal{E}$ is equivalent to $\mathcal{Z}(\overline{\mathcal{U}}) \ne 0$. Hence by (52) applied with $\delta = (c_0/2)\mathfrak{e}^{1/2}\varsigma^2$ and by the Borel-Cantelli lemma, almost surely for all large n, either $\mathcal{Z}(\overline{\mathcal{U}}) = 0$ or

$$\frac{1}{n}\log\mathcal{Z}(\overline{\mathcal{U}}) < \frac{1}{n}\mathbb{E}\left[\log\mathcal{Z}(\overline{\mathcal{U}}) \mid \mathcal{G}_t\right] + \frac{c_0}{2}\mathfrak{e}^{\frac{1}{2}}\varsigma^2 < \Psi_{\mathrm{RS}} - \frac{c_0}{2}\mathfrak{e}^{1/2}\varsigma^2.$$

Then combining with $n^{-1}\log\mathcal{Z} \to \Psi_{\mathrm{RS}}$ by Corollary 3, almost surely for all large n,

$$\langle\mathbb{I}\left(\left|\frac{1}{n}\|\sigma-\beta^\star\|^2 - 2\eta_*^{-1}\right| > \varsigma\right)\rangle = \frac{\mathcal{Z}(\overline{\mathcal{U}})}{\mathcal{Z}} < \exp\left(-\frac{c_0}{2}\mathfrak{e}^{1/2}\varsigma^2\cdot n\right).$$

Since $|n^{-1}\|\sigma-\beta^\star\|^2 - 2\eta_*^{-1}| \le 2K$, this tail bound implies

$$\left|\frac{1}{n}\langle\|\sigma-\beta^\star\|^2\rangle - 2\eta_*^{-1}\right| \le \left\langle\left|\frac{1}{n}\|\sigma-\beta^\star\|^2 - 2\eta_*^{-1}\right|\right\rangle \le \varsigma + 2K\cdot\exp\left(-\frac{c_0}{2}\mathfrak{e}^{1/2}\varsigma^2\cdot n\right) < 2\varsigma$$

almost surely for all large n. Since $\varsigma > 0$ is arbitrary, this shows $(2n)^{-1}\langle\|\sigma-\beta^\star\|^2\rangle \to \eta_*^{-1}$ almost surely. Then by Lemma 6 and dominated convergence theorem, also $\mathrm{mmse}_n = (2n)^{-1}\mathbb{E}[\langle\|\sigma-\beta^\star\|^2\rangle \mid A] \to \eta_*^{-1}$ almost surely. ∎

C. The TAP Equations

We note that the stationary initialization for VAMP in (21) requires knowledge of $\beta^\star$, and hence the resulting iterates do not define estimators of $\beta^\star$ given only (y,A). Here, we consider VAMP from the non-informative initialization $r_2^1 = 0$ and $\gamma_{2,1} = \rho_*^{-1}$. We first use Theorem 1.9 already proven to show convergence of the VAMP state evolution; a different argument for this convergence has also been given recently in [45].

Proposition 5: Consider the VAMP algorithm (18) with initialization $r_2^1 = 0$ and $\gamma_{2,1} = \rho_*^{-1}$. Under Assumptions 1.1–1.4, there exists a constant $\mathfrak{e}_0 = \mathfrak{e}_0(\mathfrak{C}) > 0$ such that if $\mathfrak{e} < \mathfrak{e}_0$, then $(\eta_{1,t})_{t\ge1}$, $(\gamma_{1,t})_{t\ge1}$, $(\eta_{2,t})_{t\ge1}$, and $(\gamma_{2,t})_{t\ge1}$ are monotone increasing and converge to $\eta_*,\gamma_*,\eta_*,\eta_*-\gamma_*$ respectively. Consequently, for $\hat\beta_j^t$ as defined in Theorem 2.1,

$$\lim_{t\to\infty}\lim_{n,m\to\infty}\frac{1}{n}\mathbb{E}[\|\hat\beta_j^t-\beta^\star\|^2 \mid A] = \eta_*^{-1}$$

where the inner limit exists almost surely.

Proof of Proposition 5: From Theorem 2.1, for both $j = 1,2$ and each fixed t, we have almost surely

$$\lim_{n,m\to\infty}\frac{1}{n}\|\hat\beta_j^t-\beta^\star\|^2 = \eta_{j,t}^{-1}.$$

To apply the dominated convergence theorem, note that under Assumption 1.2, the largest and smallest eigenvalues of $A^\top A$ converge to $d_+,d_- \ge 0$. Then applying $\gamma_{2,t} > 0$ by Theorem 2.1 and that $r \mapsto f(r,\gamma_{1,t})$ is Lipschitz by Proposition 14, from the forms of the iterations (18), there are constants $C_t > 0$ (depending on the state evolution parameters (17) and the value of $f(0,\gamma_{1,t})$) for which, for all large n,

$$\|r_1^t\| \le C_t(\|y\| + \|r_2^t\|), \qquad \|r_2^{t+1}\| \le C_t(\sqrt{n} + \|r_1^t\|).$$

Iterating these bounds and applying the definitions of $\hat\beta_j^t$, there are constants $C_t' > 0$ for which, for both $j = 1,2$ and all large n,

$$\|\hat\beta_j^t\|/\sqrt{n} \le C_t'\left(1 + (\|y\|/\sqrt{n})^t\right).$$

Then, applying $\|y\| \le \|A\|\,\|\beta^\star\| + \|\epsilon\|$, for some constants $C_t'' > 0$,

$$\frac{\|\hat\beta_j^t\|^4}{n^2} \le C_t''\left(1 + \left(\frac{\|\beta^\star\|^2}{n}\right)^{2t} + \left(\frac{\|\epsilon\|^2}{n}\right)^{2t}\right) \le C_t''\left(1 + \frac{1}{n}\sum_{i=1}^n\left(\beta_i^{\star 4t} + \epsilon_i^{4t}\right)\right). \tag{53}$$

For each fixed t, this upper bound has finite expectation independent of n, so $\{\|\hat\beta_j^t\|^2/n\}_{n\ge1}$ is bounded in L^2 and hence uniformly integrable. Then the dominated convergence theorem implies

$$\lim_{n,m\to\infty}\frac{1}{n}\mathbb{E}[\|\hat\beta_j^t-\beta^\star\|^2 \mid A] = \eta_{j,t}^{-1}.$$

Combining this with Theorem 1.9, we must have $\eta_{j,t}^{-1} \ge \eta_*^{-1}$ for every t, because each $\hat\beta_j^t$ is a (y,A)-measurable estimator of $\beta^\star$ and hence $n^{-1}\mathbb{E}[\|\hat\beta_j^t-\beta^\star\|^2 \mid A] \ge \mathrm{mmse}_n$.

It remains to show the monotonicity and convergence of $(\eta_{j,t})_{t\ge1}$ and $(\gamma_{j,t})_{t\ge1}$. Applying the definition of the

R-transform to write $\gamma_{1,t} = \eta_{2,t} - \gamma_{2,t} = -R(\eta_{2,t})$, the iterations (17) yield

$$\eta_{2,t+1}^{-1} = G\left(\text{mmse}\left(-R\left(\eta_{2,t}^{-1}\right)\right)^{-1} + R\left(\eta_{2,t}^{-1}\right)\right). \tag{54}$$

We claim that for any $x_1 \in [\eta_*^{-1}, G(-d_-))$,

$$x_2 := G\left(\text{mmse}\left(-R\left(x_1\right)\right)^{-1} + R\left(x_1\right)\right) \leq x_1 \tag{55}$$

To see this, note that $R(x)$ is negative and increasing by Lemma 13, and $\text{mmse}(\gamma)$ is decreasing over $\gamma > 0$ (by the law of total variance). Then $x \mapsto \text{mmse}(-R(x))$ is increasing over $x \in (0, G(-d_-))$. Proposition 1 implies that this function has a unique fixed point $x = \eta_*^{-1} \in (0, G(-d_-))$. Furthermore $\lim_{x\to 0} \text{mmse}(-R(x)) = \text{mmse}(1) > 0$ strictly, because the prior distribution π has strictly positive variance. Then $\text{mmse}(-R(x)) > x$ for $x < \eta_*^{-1}$ and $\text{mmse}(-R(x)) < x$ for $x > \eta_*^{-1}$, so in particular

$$\text{mmse}\left(-R\left(x_1\right)\right) \leq x_1$$

when $x_1 \geq \eta_*^{-1}$. Then $\text{mmse}(-R(x_1))^{-1} + R(x_1) \geq x_1^{-1} + R(x_1) = G^{-1}(x_1) > -d_-$. So x_2 is well-defined in (55), and also $x_2 \leq G(G^{-1}(x_1)) = x_1$ as desired because G is decreasing.

Finally, since $\eta_{2,t}^{-1} \geq \eta_*^{-1}$ for all $t \geq 1$, (55) implies that $(\eta_{2,t}^{-1})_{t\geq 1}$ is a monotonically decreasing sequence, which must then converge to a fixed point of $x \mapsto G(\text{mmse}(-R(x))^{-1} + R(x))$. Such a fixed point satisfies $G^{-1}(x) = \text{mmse}(-R(x))^{-1} + R(x)$, i.e. $x = \text{mmse}(-R(x))$, so it must be the unique fixed point η_*^{-1}. Thus $(\eta_{2,t})_{t\geq 1}$ monotonically increases to η_*. It is then straightforward to verify from their definitions in (17) that $(\gamma_{2,t})_{t\geq 1}$, $(\gamma_{1,t+1})_{t\geq 1}$, and $(\eta_{1,t+1})_{t\geq 1}$ also monotonically increase to $\eta_* - \gamma_*$, γ_*, and η_*. ∎

Proof of Theorem 1.10, Bounded Support: First note the following "Pythagorean relation": For $j = 1, 2$ and any $t \geq 1$,

$$\mathbb{E}[\|\beta^\star - \hat{\beta}_j^t\|^2 \mid A] = \mathbb{E}[\|\beta^\star - \langle\sigma\rangle\|^2 \mid A] + \mathbb{E}[\|\hat{\beta}_j^t - \langle\sigma\rangle\|^2 \mid A]$$

because $\hat{\beta}_j^t$ is a (y, A)-measurable estimator of $\beta^\star$ and $\langle\sigma\rangle = \mathbb{E}[\beta^\star \mid y, A]$. Then by Theorem 1.9 and Proposition 5, for both $j = 1, 2$,

$$\lim_{t\to\infty} \lim_{n,m\to\infty} \frac{1}{n}\mathbb{E}[\|\hat{\beta}_j^t - \langle\sigma\rangle\|^2 \mid A] = 0 \tag{56}$$

where the inner limit exists almost surely. It follows from this, uniform boundedness of $(\gamma_{2,t})_{t\geq 1}$ and $(\eta_{1,t})_{t\geq 1}$, the convergence $\eta_{1,t} - \gamma_{2,t} = \gamma_{1,t-1} \to \gamma_*$ from Proposition 5, and the triangle inequality that

$$\lim_{t\to\infty} \lim_{n,m\to\infty} \frac{1}{n}\mathbb{E}\Bigg[\bigg\| \left(A^\top A\langle\sigma\rangle - \gamma_*\langle\sigma\rangle - A^\top y\right) - \left((A^\top A + \gamma_{2,t} I)\hat{\beta}_2^t - \eta_{1,t}\hat{\beta}_1^t - A^\top y\right)\bigg\|^2 \;\bigg|\; A\Bigg] = 0. \tag{57}$$

From the definitions of $\hat{\beta}_j^t$ and (18b), $\gamma_{2,t} r_2^t = \left(A^\top A + \gamma_{2,t} I\right)\hat{\beta}_2^t - A^\top y = \eta_{1,t}\hat{\beta}_1^t - \gamma_{1,t-1} r_1^{t-1}$. Substituting this identity into (57),

$$\lim_{t\to\infty} \lim_{n,m\to\infty} \frac{1}{n}\mathbb{E}\Bigg[\bigg\| \left(A^\top A\langle\sigma\rangle - \gamma_*\langle\sigma\rangle - A^\top y\right) + \gamma_{1,t-1} r_1^{t-1}\bigg\|^2 \;\bigg|\; A\Bigg] = 0.$$

Then dividing by $\gamma_{1,t-1}^2$, applying $\gamma_{1,t-1} \to \gamma_*$, and applying that $r \mapsto f(r, \gamma_*)$ is Lipschitz by Proposition 14

$$\lim_{t\to\infty} \lim_{n,m\to\infty} \frac{1}{n}\mathbb{E}\Bigg[\bigg\| f\left(-\gamma_*^{-1}\left(A^\top A\langle\sigma\rangle - \gamma_*\langle\sigma\rangle - A^\top y\right), \gamma_*\right) - f(r_1^{t-1}, \gamma_*)\bigg\|^2 \;\bigg|\; A\Bigg] = 0. \tag{58}$$

Finally, applying again that $r \mapsto f(r, \gamma)$ is Lipschitz, (19) implies that for each fixed t we have almost surely

$$\lim_{n,m\to\infty} \frac{1}{n}\left\|f(r_1^{t-1}, \gamma_*) - f(r_1^{t-1}, \gamma_{1,t-1})\right\|^2 = \mathbb{E}\left[\left(f(\mathsf{R}_1^{t-1}, \gamma_*) - f(\mathsf{R}_1^{t-1}, \gamma_{1,t-1})\right)^2\right]$$

Here, assuming that π has bounded support, $f(r, \gamma)$ is bounded, so the dominated convergence theorem yields

$$\lim_{n,m\to\infty} \frac{1}{n}\mathbb{E}\left[\left\|f(r_1^{t-1}, \gamma_*) - f(r_1^{t-1}, \gamma_{1,t-1})\right\|^2 \;\middle|\; A\right] = \mathbb{E}\left[\left(f(\mathsf{R}_1^{t-1}, \gamma_*) - f(\mathsf{R}_1^{t-1}, \gamma_{1,t-1})\right)^2\right]. \tag{59}$$

Representing $\mathsf{R}_1^{t-1} = \mathsf{X}^\star + \mathsf{Z}/\sqrt{\gamma_{1,t-1}}$ and applying $\gamma_{1,t-1} \to \gamma_*$, an application of the dominated convergence theorem shows that the right side converges to 0 as $t \to \infty$. Then, recalling $f(r_1^{t-1}, \gamma_{1,t-1}) = \hat{\beta}_j^1$ and applying also the statement (56) for $\hat{\beta}_j^1$ and the triangle inequality, this shows

$$\lim_{t\to\infty} \lim_{n,m\to\infty} \frac{1}{n}\mathbb{E}\left[\left\|f(r_1^{t-1}, \gamma_*) - \langle\sigma\rangle\right\|^2 \;\middle|\; A\right] = 0.$$

Combining this with (58) concludes the proof. ∎

Appendix A
State Evolution of VAMP

We prove Theorems 2.1, 2.2 and Propositions 2, 3 on Vector AMP. These results do not require the high temperature condition of Assumption 1.4, and the stationary initialization (21) of VAMP and associated scalar parameters may be defined with respect to any fixed point $(\eta_*^{-1}, \gamma_*) \in (0, G(-d_-)) \times \mathbb{R}_+$ of (11).

Proof of Theorem 2.1: This follows from [21, Theorems 1 and 2]. The state evolution (17) corresponds to the setting of matched MMSE denoising described in [21, Theorem 2], where $\mathcal{E}_1(\gamma_{1,t}) = \text{mmse}(\gamma_{1,t})$ and $\mathcal{E}_2(\gamma_{2,t}) = G(\gamma_{2,t})$ as shown in [21, Eq. (41)]. (Our quantities $\eta_{j,t}, \gamma_{j,t}$ defined by (17) are the asymptotic quantities $\overline{\eta}_{jt}, \overline{\gamma}_{jt}$ in [21].) We note that under (17), $\gamma_{1,t} = \eta_{2,t} - G^{-1}(\eta_{2,t}^{-1}) = -R(\eta_{2,t}^{-1})$, which is positive for $\eta_{2,t} > 0$ by Lemma 13(b). Furthermore, setting $A(\gamma) = \gamma \cdot \text{mmse}(\gamma)$, we have $\gamma_{2,t+1} = (1 - A(\gamma_{1,t}))/\text{mmse}(\gamma_{1,t})$. The argument of [21, Section IV.F] shows $A(\gamma) \in (0, 1)$ for any $\gamma > 0$, hence $\gamma_{2,t+1} > 0$ when $\gamma_{1,t} > 0$. Thus the parameters of (17) are all positive and well-defined.

The proof of [21, Theorem 1] is easily adapted to start from an initialization $(r_2^1, \gamma_{2,1})$ instead of $(r_1^0, \gamma_{1,0})$, and to use the deterministic state evolution parameters (17) and the associated quantities $\alpha_{j,t} = \gamma_{j,t}/\eta_{j,t}$ instead of their empirical estimates. (We initialize with $(r_2^1, \gamma_{2,1})$ so that $\gamma_{2,1}^{-1}$ is correctly matched with the variance of $r_2^1 - \beta^\star$, without requiring r_1^0 to have a limit R_1^0 that is a Gaussian perturbation of $\mathsf{X}^\star$.) The empirical convergence of $D^\top 1_{m\times 1}$ and of $(r_2^1, \beta^\star)$ on Pseudo-Lipschitz test functions of order 2 (as required in [21]) are implied by our assumptions of empirical Wasserstein convergence at all orders. The first condition of [21, Theorem 1] follows from $A(\gamma) \in (0,1)$ discussed above, the second condition from the continuity of $\mathrm{mmse}(\cdot)$ and $G(\cdot)$, and the final uniform Lipschitz condition for $f(\cdot,\gamma)$ from Assumption 1.3. The convergence (19) is then shown in [21, Theorem 1, Eq. (45)], and (20) is shown in [21, Theorem 2, Eq. (56c)]. ■

A. Identities for Stationary VAMP

Proposition 6: Suppose Assumptions 1.1 and 1.3 hold, and let $(\eta_*^{-1}, \gamma_*) \in (0, G(-d_-)) \times \mathbb{R}_+$ be any fixed point of (11). Then $\eta_*^{-1} \le \rho_*$ and $\eta_* - \gamma_* \ge \rho_*^{-1} > 0$.

Proof: Let $(\mathsf{X}^\star, \mathsf{Y})$ be as defined in the scalar channel (7). The law of total variance implies $\eta_*^{-1} = \mathrm{mmse}(\gamma_*) \le \mathbb{V}(\mathsf{X}^\star) = \rho_*$. The second claim $\eta_* - \gamma_* \ge \rho_*^{-1}$ follows from comparing the MMSE with the error of the linear estimator $a\mathsf{Y}$ with $a = \frac{\rho_*}{\rho_* + \gamma_*^{-1}}$:

$$\eta_*^{-1} = \mathbb{E}\left(\mathsf{X}^\star - \mathbb{E}\left[\mathsf{X}^\star \mid \mathsf{Y}\right]\right)^2 \le \mathbb{E}\left(\mathsf{X}^\star - a\left(\mathsf{X}^\star + \frac{1}{\sqrt{\gamma_*}}\mathsf{Z}\right)\right)^2$$
$$= (1-a)^2\rho_* + \frac{a^2}{\gamma_*} = \frac{1}{\rho_*^{-1} + \gamma_*}. \tag{60}$$

Rearranging yields $\eta_* - \gamma_* \ge \rho_*^{-1}$. ■

Proof of Proposition 2: Writing (11) as $\eta_*^{-1} = \mathrm{mmse}(\gamma_*)$ and $\eta_*^{-1} = G(\eta_* - \gamma_*)$, it is clear that the initialization $\gamma_{1,0} = \gamma_*$ yields $\eta_{1,t} = \eta_{2,t} = \eta_*$, $\gamma_{1,t} = \gamma_*$, and $\gamma_{2,t} = \eta_* - \gamma_*$ for all t.

Then substituting $y = A\beta^\star + \epsilon$ and $A = Q^\top DO$, the update rule (18a) for r_1^t can be rearranged as

$$r_1^t = O^\top \Lambda O r_2^t + O^\top\left[\frac{\eta_*}{\gamma_*}(D^\top D + (\eta_* - \gamma_*)I)^{-1}D^\top D\right]O\beta^\star + e.$$

Applying the identity $(D^\top D + (\eta_* - \gamma_*)I)^{-1}D^\top D = I - (\eta_* - \gamma_*)(D^\top D + (\eta_* - \gamma_*)I)^{-1} = (\gamma_*/\eta_*)(I - \Lambda)$ we obtain

$$y^t := r_1^t - e - \beta^\star = O^\top \Lambda O(x^t + \beta^\star) + O^\top(I - \Lambda)O\beta^\star - \beta^\star$$
$$= O^\top \Lambda O x^t$$

which may be written as $s^t = Ox^t$ and $y^t = O^\top \Lambda s^t$. Setting $r_1^t = p^t + \beta^\star$, we have from (18b)

$$x^{t+1} := r_2^{t+1} - \beta^\star$$
$$= \frac{\eta_*}{\eta_* - \gamma_*}f(p^t + \beta^\star, \gamma_{1,t}) - \frac{\gamma_*}{\eta_* - \gamma_*}(p^t + \beta^\star) - \beta^\star$$
$$= F(p^t, \beta^\star).$$

For $t = 0$, this gives the initialization $x^1 = F(p^0, \beta^\star)$, and for $t \ge 1$, we have $p^t = y^t + e$ so this gives the update for x^{t+1}. ■

Proposition 7: Suppose Assumptions 1.1 and 1.3 hold, and let (η_*^{-1}, γ_*) be any fixed point of (11). Let Y denote an observation from the scalar channel (7) with variance γ_*^{-1}. Then the functions $y \mapsto f(y, \gamma_*)$ and $(p, \beta) \mapsto F(p, \beta)$ are continuously differentiable and Lipschitz. We have

$$f'(y, \gamma_*) := \frac{\partial}{\partial y}f(y, \gamma_*) = \gamma_* \mathbb{V}(\mathsf{X}^\star \mid \mathsf{Y} = y),$$
$$F'(p, \beta) := \frac{\partial}{\partial p}F(p, \beta) = \frac{\eta_*}{\eta_* - \gamma_*}f'(p + \beta, \gamma_*) - \frac{\gamma_*}{\eta_* - \gamma_*}, \tag{61}$$

and these are non-constant in y and p. For $\mathsf{P} \sim N(0, \gamma_*^{-1})$ and $\mathsf{X}^\star \sim \pi$ independent, we also have

$$\mathbb{E}F(\mathsf{P}, \mathsf{X}^\star) = 0, \quad \mathbb{E}F'(\mathsf{P}, \mathsf{X}^\star) = 0, \quad \mathbb{E}\left(F(\mathsf{P}, \mathsf{X}^\star)^2\right) = \delta_* \tag{62}$$

Proof: Proposition 14 shows that $y \mapsto f(y, \gamma_*)$ is continuously-differentiable and Lipschitz, with derivative given by (61). Since π is also non-Gaussian by Assumption 1.3, $y \mapsto f(y, \gamma_*)$ is non-linear, and hence $y \mapsto f'(y, \gamma_*)$ is non-constant. Then the same properties hold for F and F', and the form (61) for F' follows from definition of F.

The first two identities in (62) follow from $\mathbb{E}\mathsf{P} = \mathbb{E}\mathsf{X}^\star = \mathbb{E}f(\mathsf{P} + \mathsf{X}^\star, \gamma_*) = 0$ and $\mathbb{E}f'(\mathsf{P} + \mathsf{X}^\star, \gamma_*) = \gamma_* \cdot \mathrm{mmse}(\gamma_*) = \gamma_*/\eta_*$. For the last identity in (62), denote for simplicity $f(y) = f(y, \gamma_*)$. Note that

$$\mathbb{E}\left(f(\mathsf{P} + \mathsf{X}^\star) - \mathsf{X}^\star\right)^2 = \mathrm{mmse}(\gamma_*) = \eta_*^{-1}$$

and thus

$$\left(\frac{\eta_*}{\eta_* - \gamma_*}\right)^2\left(\mathbb{E}\left(f(\mathsf{P} + \mathsf{X}^\star) - \mathsf{X}^\star\right)^2 - \frac{\gamma_*}{\eta_*^2}\right)$$
$$= \left(\frac{\eta_*}{\eta_* - \gamma_*}\right)^2\left(\frac{1}{\eta_*} - \frac{\gamma_*}{\eta_*^2}\right)$$
$$= \frac{1}{\eta_* - \gamma_*} = \delta_*. \tag{63}$$

It follows that

$$\mathbb{E}\left(F(\mathsf{P}, \mathsf{X}^\star)^2\right)$$
$$= \mathbb{E}\left(\frac{\eta_*}{\eta_* - \gamma_*}f(\mathsf{P} + \mathsf{X}^\star) - \frac{\gamma_*}{\eta_* - \gamma_*}\mathsf{P} - \frac{\eta_*}{\eta_* - \gamma_*}\mathsf{X}^\star\right)^2$$
$$= \left(\frac{\eta_*}{\eta_* - \gamma_*}\right)^2\left(\mathbb{E}\left(f(\mathsf{P} + \mathsf{X}^\star) - \mathsf{X}^\star\right)^2 + \left(\frac{\gamma_*}{\eta_*}\right)^2\mathbb{E}\mathsf{P}^2 - \frac{2\gamma_*}{\eta_*}\mathbb{E}\left(\left(f(\mathsf{P} + \mathsf{X}^\star) - \mathsf{X}^\star\right)\mathsf{P}\right)\right)$$
$$\overset{(a)}{=} \left(\frac{\eta_*}{\eta_* - \gamma_*}\right)^2\left(\mathbb{E}\left(f(\mathsf{P} + \mathsf{X}^\star) - \mathsf{X}^\star\right)^2 + \frac{\gamma_*}{\eta_*^2} - \frac{2}{\eta_*}\mathbb{E}\left(f'(\mathsf{P} + \mathsf{X}^\star)\right)\right)$$
$$\overset{(b)}{=} \left(\frac{\eta_*}{\eta_* - \gamma_*}\right)^2\mathbb{E}\left(\left(f(\mathsf{P} + \mathsf{X}^\star) - \mathsf{X}^\star\right)^2 - \frac{\gamma_*}{\eta_*^2}\right) = \delta_*$$

where we used Gaussian integration by parts in (a) and $\mathbb{E}f'(\mathsf{P} + \mathsf{X}^\star, \gamma_*) = \gamma_*/\eta_*$ and (63) in (b). ■

Lemma 7: Recall b_*, κ_* from (25) and a_*, c_*, e_* from (31). We have the following identities

$$d_*^A := \mathbb{E}\frac{1}{\mathsf{D}^2+\eta_*-\gamma_*} = \frac{1}{\eta_*} \tag{64a}$$

$$d_*^B := \mathbb{E}\frac{\mathsf{D}^2}{\mathsf{D}^2+\eta_*-\gamma_*} = \frac{\gamma_*}{\eta_*} \tag{64b}$$

$$d_*^C := \mathbb{E}\frac{1}{(\mathsf{D}^2+\eta_*-\gamma_*)^2} = \frac{1}{\eta_*^2}\left(\frac{\gamma_*}{\eta_*-\gamma_*}\right)^2\kappa_* + \frac{1}{\eta_*^2} \tag{64c}$$

$$d_*^D := \mathbb{E}\frac{\mathsf{D}^2}{(\mathsf{D}^2+\eta_*-\gamma_*)^2} = -\frac{\gamma_*^2}{\eta_*^2}\left(\frac{\kappa_*}{\eta_*-\gamma_*}\right) + \frac{\gamma_*}{\eta_*^2} \tag{64d}$$

$$d_*^E := \mathbb{E}\frac{\mathsf{D}^4}{(\mathsf{D}^2+\eta_*-\gamma_*)^2} = \frac{\gamma_*^2}{\eta_*^2}(1+\kappa_*) \tag{64e}$$

Also, for the quantities L, E_b from Proposition 3, we have

$$\mathbb{E}\mathsf{L} = 0, \quad \mathbb{E}\mathsf{L}^2 = \kappa_*, \quad \mathbb{E}\mathsf{E}_b^2 = b_*,$$
$$\mathbb{E}\mathsf{D}^2\mathsf{L} = a_*, \quad \mathbb{E}\mathsf{D}^2\mathsf{L}^2 = c_*, \quad \mathbb{E}\mathsf{D}^2\mathsf{E}_b^2 = e_*. \tag{65}$$

Proof: (64a) is the identity $\eta_*^{-1} = G(\eta_*-\gamma_*)$, which is a rewriting of the second equation of (11). This then implies $\mathbb{E}\mathsf{L} = 0$, as well as $\mathbb{E}\mathsf{L}^2 = \kappa_*$. (64b) follows from the identity $d_*^B + (\eta_*-\gamma_*)\,d_*^A = 1$, and this then implies $\mathbb{E}\mathsf{D}^2\mathsf{L} = a_*$. (64c) follows from rearranging the identity $\kappa_* = \mathbb{E}\mathsf{L}^2 = \frac{(\eta_*-\gamma_*)^2}{\gamma_*^2}(\eta_*^2 d_*^C - 1)$. (64d) then follows from the identity $d_*^D + (\eta_*-\gamma_*)\,d_*^C = d_*^A$, and this then implies $\mathbb{E}\mathsf{E}_b^2 = b_*$ as well as $\mathbb{E}\mathsf{D}^2\mathsf{L}^2 = c_*$. Finally, (64e) follows from the identity $d_*^E + (\eta_*-\gamma_*)^2\,d_*^C + 2\,(\eta_*-\gamma_*)\,d_*^D = 1$, and this then implies $\mathbb{E}\mathsf{D}^2\mathsf{E}_b^2 = e_*$. ■

Remark 1.1: This shows $b_*, \kappa_* > 0$ strictly, because D has strictly positive variance by Theorem 1.1, and hence so do L and E_b. Also by Proposition 6, we have $\eta_* - \gamma_* > 0$, hence $\delta_*, \sigma_*^2 > 0$.

B. State Evolution of Stationary VAMP

Proof of Proposition 3: Note that $\xi = Q\epsilon \sim N(0, I_{m\times m})$. Then $D^\top \xi \in \mathbb{R}^n$ may be written as the entrywise product of $D^\top 1_{m\times 1} \in \mathbb{R}^n$ and a vector $\bar{\xi} \sim N(0, I_{n\times n})$, both when $m \ge n$ and when $m \le n$. The almost-sure convergence $H \overset{W}{\to} \mathsf{H}$ is then a straightforward consequence of Propositions 9, 10, and 12, where all random variables of H have finite moments of all orders under Assumptions 1.2 and 1.3. The identities $\kappa_* = \mathbb{E}\mathsf{L}^2$ and $b_* = \mathbb{E}\mathsf{E}_b^2$ were shown in (65). ■

Proof of Theorem 2.2: We have $\delta_{11} = \mathbb{E}\mathsf{X}_1^2 = \delta_*$ by the last identity of (62). Supposing that $\delta_{tt} = \mathbb{E}\mathsf{X}_t^2 = \delta_*$, we have by definition $\mathbb{E}\mathsf{Y}_t^2 = \kappa_*\delta_{tt} = \sigma_*^2 = \delta_*\kappa_*$. Since Y_t is independent of E, we have $\mathsf{Y}_t + \mathsf{E} \sim N(0, \sigma_*^2 + b_*)$ where this variance is $\sigma_*^2 + b_* = \gamma_*^{-1}$ by the definition of b_*. Then $\mathbb{E}\mathsf{X}_{t+1}^2 = \delta_*$ by the last identity of (62), so $\mathbb{E}\mathsf{X}_t^2 = \delta_*$ and $\mathbb{E}\mathsf{Y}_t^2 = \sigma_*^2$ for all $t \ge 1$.

Noting that Δ_t is the upper-left submatrix of Δ_{t+1}, let us denote

$$\Delta_{t+1} = \begin{pmatrix} \Delta_t & \delta_t \\ \delta_t^\top & \delta_* \end{pmatrix}$$

We now show by induction on t the following three statements:

1) $\Delta_t \succ 0$ strictly.

2) We have

$$\mathsf{Y}_t = \sum_{k=1}^{t-1}\mathsf{Y}_k\left(\Delta_{t-1}^{-1}\delta_{t-1}\right)_k + \mathsf{U}_t,$$
$$\mathsf{S}_t = \sum_{k=1}^{t-1}\mathsf{S}_k\left(\Delta_{t-1}^{-1}\delta_{t-1}\right)_k + \mathsf{U}_t' \tag{66}$$

where $\mathsf{U}_t, \mathsf{U}_t'$ are Gaussian variables with strictly positive variance, independent of H, $(\mathsf{Y}_1, \ldots, \mathsf{Y}_{t-1})$, and $(\mathsf{S}_1, \ldots, \mathsf{S}_{t-1})$.

3) We have

$$(H, X_{t+1}, S_t, Y_t) \overset{W}{\to} (\mathsf{H}, \mathsf{X}_1, \ldots, \mathsf{X}_{t+1}, \mathsf{S}_1, \ldots, \mathsf{S}_t, \mathsf{Y}_1, \ldots, \mathsf{Y}_t).$$

We take as base case $t = 0$, where the first two statements are vacuous, and the third statement requires $(H, x^1) \overset{W}{\to} (\mathsf{H}, \mathsf{X}_1)$ almost surely as $n \to \infty$. Recall that $x^1 = F(p^0, \beta^\star)$, and that $F(p, \beta)$ is Lipschitz by Proposition 7. Then this third statement follows from Propositions 3 and 10.

Supposing that these statements hold for some $t \ge 0$, we now show that they hold for $t+1$. To show the first statement $\Delta_{t+1} \succ 0$, note that for $t = 0$ this follows from $\Delta_1 = \delta_* > 0$ by Remark 1.1. For $t \ge 1$, given that $\Delta_t \succ 0$, Δ_{t+1} is singular if and only if there exist constants $\alpha_1, \ldots, \alpha_t \in \mathbb{R}$ such that

$$\mathsf{X}_{t+1} = F(\mathsf{Y}_t + \mathsf{E}, \mathsf{X}^\star) = \sum_{r=1}^{t}\alpha_r\mathsf{X}_r$$

with probability 1. From the induction hypothesis, $\mathsf{Y}_t = \sum_{k=1}^{t-1}\mathsf{Y}_k\left(\Delta_r^{-1}\delta_r\right)_k + \mathsf{U}_t$ where U_t is independent of $\mathsf{H}, \mathsf{Y}_1, \ldots, \mathsf{Y}_{t-1}$ and hence also of $\mathsf{E}, \mathsf{X}^\star, \mathsf{X}_1, \ldots, \mathsf{X}_t$. We now show that for any realized values (e_0, x_0, w_0) of

$$\left(\mathsf{E} + \sum_{k=1}^{t-1}\mathsf{Y}_k\left(\Delta_r^{-1}\delta_r\right)_k, \quad \mathsf{X}^\star, \quad \sum_{r=1}^{t}\alpha_r\mathsf{X}_r\right),$$

we have that $\mathbb{P}\left(F\left(\mathsf{U}_t + e_0, x_0\right) \ne w_0\right) > 0$. This would imply that $\Delta_{t+1} \succ 0$. From Proposition 7, $f'(y, \gamma_*)$ is non-constant, so there exists $y \in \mathbb{R}$ such that $f'(y, \gamma_*) \ne \gamma_*/\eta_*$. This implies that there exists some $u_0 \in \mathbb{R}$ such that

$$F'\left(u_0 + e_0, x_0\right) = \frac{\eta_*}{\eta_*-\gamma_*}f'\left(u_0 + e_0 + x_0, \gamma_*\right) - \frac{\gamma_*}{\eta_*-\gamma_*} \ne 0.$$

Then by the inverse function theorem, $F(u + e_0, x_0) = w_0$ has at most one solution for u in an open neighborhood of u_0. Since U_t is Gaussian with strictly positive variance, this shows $\mathbb{P}\left(F\left(\mathsf{U}_t + e_0, x_0\right) \ne w_0\right) > 0$ as desired. We thus have proved the first inductive statement that $\Delta_{t+1} \succ 0$.

To study the empirical limit of s_{t+1}, let $U = (e_b, S_t, \Lambda S_t)$ and $V = (e, X_t, Y_t)$. (For $t = 0$, this is simply $U = e_b$ and $V = e$.) By the induction hypothesis, the independence of $(\mathsf{S}_1, \ldots, \mathsf{S}_t)$ with $(\mathsf{E}_b, \mathsf{L})$, and the identities $\mathbb{E}\mathsf{E}_b^2 = b_*$ and $\mathbb{E}\mathsf{L} = 0$ and $\mathbb{E}\mathsf{L}^2 = \kappa_*$, almost surely as $n \to \infty$,

$$\frac{1}{n}\left(e_b, S_t, \Lambda S_t\right)^\top\left(e_b, S_t, \Lambda S_t\right) \to \begin{pmatrix} b_* & 0 & 0 \\ 0 & \Delta_t & 0 \\ 0 & 0 & \kappa_*\Delta_t \end{pmatrix} \succ 0$$

So almost surely for sufficiently large n, conditional on (H, X_{t+1}, S_t, Y_t), the law of s^{t+1} is given by its law conditioned on $U = OV$, which is (see [34, Lemma B.2])

$$s^{t+1}\big|_{U=OV} = Ox^{t+1}\big|_{U=OV} \stackrel{L}{=} U\left(U^\top U\right)^{-1} V^\top x^{t+1} + \Pi_{U^\perp}\tilde{O}\Pi_{V^\perp}^\top x^{t+1} \quad (67)$$

where $\tilde{O} \sim \text{Haar}(\mathbb{SO}(n-(2t+1)))$ and $\Pi_{U^\perp}, \Pi_{V^\perp} \in \mathbb{R}^{n\times(n-(2t+1))}$ are matrices with orthonormal columns spanning the orthogonal complements of the column spans of U, V respectively. We may replace s^{t+1} by the right side of (67) without affecting the joint law of $\left(H, X_{t+1}, S_t, Y_t, s^{t+1}\right)$.

For $t = 0$, we have $\mathbb{E}\mathsf{X}_1\mathsf{E} = 0$ since X_1 is independent of E. For $t \geq 1$, by the definition of X_{t+1}, the condition $\mathbb{E}F'(\mathsf{P}, \mathsf{X}^\star) = 0$ from (62), and Stein's lemma, we have $\mathbb{E}\mathsf{X}_{t+1}\mathsf{E} = 0$ and $\mathbb{E}\mathsf{X}_{t+1}\mathsf{Y}_r = 0$ for each $r = 1, \ldots, t$. Then by the induction hypothesis, almost surely as $n \to \infty$,

$$\left(n^{-1}U^\top U\right)^{-1} \to \begin{pmatrix} b_* & 0 & 0 \\ 0 & \Delta_t & 0 \\ 0 & 0 & \kappa_*\Delta_t \end{pmatrix}^{-1},$$

$$n^{-1}V^\top x_{t+1} \to \begin{pmatrix} 0 \\ \delta_t \\ 0 \end{pmatrix}.$$

Then by (67) and Propositions 11 and 12, it follows that

$$\left(H, X_{t+1}, S_t, Y_t, s^{t+1}\right) \xrightarrow{W} \Big(\mathsf{H}, \mathsf{X}_1, \ldots, \mathsf{X}_{t+1}, \mathsf{S}_1, \ldots, \mathsf{S}_t, \mathsf{Y}_1, \ldots \mathsf{Y}_t, \sum_{r=1}^t \mathsf{S}_r\left(\Delta_t^{-1}\delta_t\right)_r + \mathsf{U}'_{t+1}\Big)$$

where U'_{t+1} is the Gaussian limit of the second term on the right side of (67) and is independent of $\mathsf{H}, \mathsf{X}_1, \ldots, \mathsf{X}_{t+1}, \mathsf{S}_1, \ldots, \mathsf{S}_t, \mathsf{Y}_1, \ldots \mathsf{Y}_t$. We can thus set $\mathsf{S}_{t+1} := \sum_{r=1}^t \mathsf{S}_r\left(\Delta_t^{-1}\delta_t\right)_r + \mathsf{U}'_{t+1}$. Then $(\mathsf{S}_1, \ldots, \mathsf{S}_{t+1})$ is multivariate Gaussian and remains independent of H and $(\mathsf{Y}_1, \ldots, \mathsf{Y}_t)$. Since $n^{-1}\|s^{t+1}\|^2 = n^{-1}\|x^{t+1}\|^2 \to \delta_*$ almost surely as $n \to \infty$ by the induction hypothesis, we have $\mathbb{E}\mathsf{S}_{t+1}^2 = \delta_*$. From the form of S_{t+1}, we may check also $\mathbb{E}\mathsf{S}_{t+1}(\mathsf{S}_1, \ldots, \mathsf{S}_t) = \delta_t$, so $(\mathsf{S}_1, \ldots, \mathsf{S}_{t+1})$ has covariance Δ_{t+1} as desired. Furthermore $\sum_{r=1}^t \mathsf{S}_r\left(\Delta_t^{-1}\delta_t\right)_r \sim N\left(0, \delta_t^\top\Delta_t^{-1}\delta_t\right)$. From $\Delta_{t+1} \succ 0$ and the Schur complement formula, $\delta_* - \delta_t^\top\Delta_t^{-1}\delta_t > 0$ strictly. Then U'_{t+1} has strictly positive variance, since the variance of $\sum_{r=1}^t \mathsf{S}_r\left(\Delta_t^{-1}\delta_t\right)_r$ is less than the variance of S_{t+1}. This proves the second equation in (66) for $t+1$.

Now, we study the empirical limit of y^{t+1}. Let $U = (e, X_{t+1}, Y_t)$, $V = (e_b, S_{t+1}, \Lambda S_t)$. Similarly by the induction hypothesis and the empirical convergence of (H, S_{t+1}) already shown, almost surely as $n \to \infty$,

$$\frac{1}{n}\left(e_b, S_{t+1}, \Lambda S_t\right)^\top\left(e_b, S_{t+1}, \Lambda S_t\right) \to \begin{pmatrix} b_* & 0 & 0 \\ 0 & \Delta_{t+1} & 0 \\ 0 & 0 & \kappa_*\Delta_t \end{pmatrix} \succ 0.$$

Then the law of y^{t+1} conditional on $(H, X_{t+1}, S_{t+1}, Y_t)$ is given by its law conditioned on $U = O^\top V$, which is

$$y^{t+1}\big|_{U=O^\top V} = O^\top\Lambda s^{t+1}\big|_{U=O^\top V} \stackrel{L}{=} U\left(V^\top V\right)^{-1} V^\top\Lambda s^{t+1} + \Pi_{U^\perp}\tilde{O}\Pi_{V^\perp}^\top\Lambda s^{t+1} \quad (68)$$

where $\tilde{O} \sim \text{Haar}(\mathbb{SO}(n-(2t+2)))$. From the convergence of (H, S_{t+1}) already shown, almost surely as $n \to \infty$,

$$\left(n^{-1}V^\top V\right)^{-1} \to \begin{pmatrix} b_* & 0 & 0 \\ 0 & \Delta_{t+1} & 0 \\ 0 & 0 & \kappa_*\Delta_t \end{pmatrix}^{-1},$$

$$n^{-1}V^\top\Lambda s_{t+1} \to \begin{pmatrix} 0 \\ 0 \\ \kappa_*\delta_t \end{pmatrix}.$$

Then by (68) and Propositions 11 and 12,

$$\left(H, X_{t+1}, S_{t+1}, Y_t, y^{t+1}\right) \xrightarrow{W} \Big(\mathsf{H}, \mathsf{X}_1, \ldots, \mathsf{X}_{t+1}, \mathsf{S}_1, \ldots, \mathsf{S}_{t+1}, \mathsf{Y}_1, \ldots \mathsf{Y}_t, \sum_{r=1}^t \mathsf{Y}_r\left(\Delta_t^{-1}\delta_t\right)_r + \mathsf{U}_{t+1}\Big)$$

where U_{t+1} is the limit of the second term on the right side of (68), which is Gaussian and independent of $\mathsf{H}, \mathsf{S}_1, \ldots, \mathsf{S}_{t+1}, \mathsf{Y}_1, \ldots \mathsf{Y}_t$. Setting $\mathsf{Y}_{t+1} := \sum_{r=1}^t \mathsf{Y}_r\left(\Delta_t^{-1}\delta_t\right)_r + \mathsf{U}_{t+1}$, it follows that $(\mathsf{Y}_1, \ldots, \mathsf{Y}_{t+1})$ remains independent of H and $(\mathsf{S}_1, \ldots, \mathsf{S}_{t+1})$. We may check that $\mathbb{E}\mathsf{Y}_{t+1}(\mathsf{Y}_1, \ldots, \mathsf{Y}_t) = \kappa_*\delta_t$, and we have also $n^{-1}\|y^{t+1}\|^2 = n^{-1}\|\Lambda s^{t+1}\|^2 \to \kappa_*\delta_*$ so $\mathbb{E}\mathsf{Y}_{t+1}^2 = \kappa_*\delta_*$. From $\Delta_{t+1} \succ 0$ and the Schur complement formula, note that $\sum_{r=1}^t \mathsf{Y}_r\left(\Delta_t^{-1}\delta_t\right)_r$ has variance $\kappa_*\delta_t^\top\Delta_t^{-1}\delta_t$ which is strictly smaller than $\kappa_*\delta_*$, so U_{t+1} has strictly positive variance. This proves the first equation in (66) for $t+1$, and completes the proof of this second inductive statement.

Finally, recall $x^{t+2} = F\left(y^{t+1} + e, \beta^\star\right)$ where F is Lipschitz. Then by Proposition 10, almost surely

$$(H, X_{t+2}, S_{t+1}, Y_{t+1}) \xrightarrow{W} (\mathsf{H}, \mathsf{X}_1, \ldots, \mathsf{X}_{t+2}, \mathsf{S}_1, \ldots, \mathsf{S}_{t+1}, \mathsf{Y}_1, \ldots, \mathsf{Y}_{t+1})$$

where $\mathsf{X}_{t+2} = F(\mathsf{Y}_{t+1} + \mathsf{E}, \mathsf{X}^\star)$, showing the third inductive statement and completing the induction. ∎

Proof of Corollary 1: This follows from the empirical Wasserstein convergence of (e, X_t, Y_t) guaranteed by Theorem 2.2. The statements $n^{-1}X^\top e \to 0$ and $n^{-1}X^\top Y \to 0$ follow from the identity $\mathbb{E}F'(\mathsf{P}, \mathsf{X}^\star) = 0$ in (62) and Stein's lemma, and the remaining statements follow directly from the independence of $(\mathsf{Y}_1, \ldots, \mathsf{Y}_t)$ with E and from their specified Gaussian laws. ∎

Proof of Corollary 2: This follows directly from Theorem 2.2, the independence of $(\mathsf{S}_1, \ldots, \mathsf{S}_t)$ and D, and our definition of empirical Wasserstein convergence. ∎

Appendix B
Properties in High Temperature

We show uniqueness of the fixed point to (11) and convergence of the stationary VAMP state evolution, assuming the high temperature condition of Assumption 1.4.

A. Fixed-Point Equation

Proof: [Proof of Proposition 1] Provided that the fixed point (η_*^{-1}, γ_*) is unique, the statements $\eta_*^{-1} \le \rho_*$ and $\eta_* - \gamma_* \ge \rho_*^{-1}$ were shown in Proposition 6.

To show uniqueness of this fixed point, by the law of total variance, $\mathrm{mmse}(\gamma) \le \rho_* \le \mathfrak{C}$ for any $\gamma > 0$. Then for all $\mathfrak{e} < 1/(2\mathfrak{C})$, we have $\rho_* < G(-d_-)$ from Lemma 13(a), and $-R(\eta^{-1}) > 0$ for all $\eta^{-1} \in (0, \rho_*]$ from Lemma 13(b). Extending $-R$ by continuity to $-R(0) = -\mathbb{E}[-\mathsf{D}^2] = d_* > 0$ via (219), this shows that

$$h(\eta^{-1}) := \mathrm{mmse}(-R(\eta^{-1})) \tag{69}$$

is a well-defined continuous map from $[0, \rho_*]$ to itself. Applying Stein's lemma, the derivative of $\gamma \mapsto \mathrm{mmse}(\gamma)$ may be computed to be

$$\mathrm{mmse}'(\gamma) = -\mathbb{E}[\mathbb{V}[\mathsf{X}^\star \mid \mathsf{Y}]^2], \tag{70}$$

see e.g. [46, Theorem 2] whose scalar specialization ($m = 1$ and $G = \sqrt{\gamma}$) yields $\frac{d}{d\sqrt{\gamma}}\mathrm{mmse}(\gamma) = -2\sqrt{\gamma}\,\mathbb{E}[\mathbb{V}[\mathsf{X}^\star \mid \mathsf{Y}]^2]$, and (70) then follows from the chain rule. Then the map (69) has derivative

$$h'(\eta^{-1}) = \mathbb{E}[\mathbb{V}[\mathsf{X}^\star \mid \mathsf{Y}]^2] \cdot R'(\eta^{-1}) \tag{71}$$

where $\mathbb{E}, \mathbb{V}$ are with respect to the scalar channel (7) with inverse-variance $\gamma = -R(\eta^{-1})$. By the condition (6) of Assumption 1.3, for any such channel, $\mathbb{E}[\mathbb{V}[\mathsf{X}^\star \mid \mathsf{Y}]^2] \le C$ where $C > 0$ depends only on $\mathfrak{C}$. By Lemma 13(b–c), $R'(\eta^{-1}) > 0$ and $R'(\eta^{-1}) < O(\mathfrak{e}^2)$ for all $\eta^{-1} \in (0, \rho_*)$. Then for $\mathfrak{e} < \mathfrak{e}_0$ small enough, we have $\sup_{\eta^{-1} \in (0,\rho_*)} h'(\eta^{-1}) \in (0, 1)$ strictly. Then $h(\cdot)$ defines a contractive map on $[0, \rho_*]$, so it has a unique fixed point $\eta_*^{-1} \in [0, \rho_*]$ by the Banach fixed-point theorem. We have $h(0) = \mathrm{mmse}(d_*) > 0$ strictly, because π has strictly positive variance. Thus there is a unique fixed point (η_*^{-1}, γ_*) to (11) where $\eta_*^{-1} \in (0, \rho_*] \subset (0, G(-d_-))$ and $\gamma_* = -R(\eta_*^{-1}) > 0$. ■

B. Explicit High Temperature Condition

Proof of Theorem 1.11: Suppose Assumptions 1.1–1.4 hold for the linear model in (1). Consider the rescaled problem: $y = A_{\text{rescaled}}\beta^\star_{\text{rescaled}} + \epsilon$ where $A_{\text{rescaled}} := \sqrt{\mathfrak{C}}A$ and $\beta^\star_{\text{rescaled}} := \frac{1}{\sqrt{\mathfrak{C}}}\beta^\star$. Note that $A_{\text{rescaled}} = Q^\top D_{\text{rescaled}} O$ with $D_{\text{rescaled}} := \sqrt{\mathfrak{C}}D$ and $D_{\text{rescaled}}^\top 1_{m\times 1} \xrightarrow{W} \mathsf{D}_{\text{rescaled}}$ with $\mathsf{D}_{\text{rescaled}} := \sqrt{\mathfrak{C}}\mathsf{D}$.

We then have that (i) Assumptions 1.1 and 1.2 hold for the rescaled problem with D_{rescaled} and $\mathsf{D}_{\text{rescaled}}$ in place of D and D; (ii) Assumption 1.4 holds for the rescaled problem: $\mathrm{supp}(\mathsf{D}^2_{\text{rescaled}}) \subseteq [d_{*,\text{rescaled}} - \mathfrak{e}_{\text{rescaled}}, d_{*,\text{rescaled}} + \mathfrak{e}_{\text{rescaled}}]$ with $d_{*,\text{rescaled}} := \mathfrak{C}d_*$ and $\mathfrak{e}_{\text{rescaled}} := \mathfrak{C}\mathfrak{e}$; (iii) Assumption 1.3 holds for the rescaled problem with parameter $\mathfrak{C}_{\text{rescaled}} = 1$ in place of $\mathfrak{C}$. This follows from Proposition 13 and that the rescaled prior $\mathsf{X}^\star_{\text{rescaled}} := \frac{1}{\sqrt{\mathfrak{C}}}\mathsf{X}^\star$ either has bounded support contained in $[-1, 1]$ or admits a density function $\sqrt{\mathfrak{C}}\exp\Big\{-g(\sqrt{\mathfrak{C}}x)\Big\}$ with $\frac{d^2}{dx^2}g(\sqrt{\mathfrak{C}}x) \ge 1$.

Apply Theorem 1.7–Theorem 1.10 to the rescaled problem. Along with (ii), (iii) above, we obtain that there exists an absolute constant $\mathfrak{a} := \mathfrak{e}_0(\mathfrak{C}_{\text{rescaled}} = 1) > 0$ such that if $\mathfrak{e}_{\text{rescaled}} := \mathfrak{C}\mathfrak{e} \le \mathfrak{a}$, (13), (14) and (15) hold for the rescaled problem. It is straightforward to show that (13), (14) and (15) hold for the original problem if and only if they hold for the rescaled problem. The proof is now finished. ■

C. Scalar Parameters

Let us record here the leading-order behaviors of several quantities related to the scalar parameters of (25) and (31) in the small parameter $\mathfrak{e}$ of Assumption 1.4.

Proposition 8: Suppose Assumptions 1.1, 1.3, and 1.4 hold. Let $\kappa_2 = \mathbb{V}(\mathsf{D}^2)$ and let $R(z)$ be the R-transform of $-\mathsf{D}^2$. For some constant $\mathfrak{e}_0 = \mathfrak{e}_0(\mathfrak{C}) > 0$, if $\mathfrak{e} < \mathfrak{e}_0$, then for any constant $c > 0$ and all $z \in (0, c)$,

$$\begin{aligned} &R(z) = -d_* + \kappa_2 z(1 + z \cdot O(\mathfrak{e})), \quad R'(z) = \kappa_2 + O(\mathfrak{e}^3), \\ &R''(z) = O(\mathfrak{e}^3). \end{aligned} \tag{72}$$

Furthermore

$$\begin{aligned} &\kappa_2 \le \min(d_*\mathfrak{e}, \mathfrak{e}^2), \quad \gamma_* = d_* - \kappa_2\eta_*^{-1}\left(1 + \eta_*^{-1} \cdot O(\mathfrak{e})\right), \\ &\kappa_* = \left(\frac{\eta_* - \gamma_*}{\eta_*}\right)^2 \frac{\kappa_2}{d_*^2}\big(1 + \eta_*^{-1} \cdot O(\mathfrak{e})\big), \\ &\alpha_*^A = \frac{\eta_*}{\sqrt{\kappa_2}}\left(1 + \eta_*^{-1} \cdot O(\mathfrak{e})\right), \quad \alpha_*^B = -\eta_* + O(\mathfrak{e}), \\ &b_* = \frac{1}{d_*}\left(1 + \eta_*^{-1} \cdot O\left(\frac{\kappa_2}{d_*}\right)\right), \\ &\frac{e_*}{b_*} = d_* + O(\mathfrak{e}), \quad \frac{a_*}{\sqrt{\kappa_*}} = O(\mathfrak{e}), \quad \frac{c_*}{\kappa_*} = d_* + O(\mathfrak{e}). \end{aligned} \tag{73}$$

Proof: (72) and the bounds for κ_2 follow from Lemma 13(c) and $\mathbb{E}[-\mathsf{D}^2] = -d_*$. We will use implicitly the bound $\eta_*^{-1} \le \rho_* \le \mathfrak{C}$ from Proposition 1, and hence $\eta_*^{-1} = O(1)$, throughout the proof.

Applying (72) to the fixed point equation $\gamma_* = -R(\eta_*^{-1})$ in (11), we have

$$\gamma_* = d_* - \kappa_2\eta_*^{-1}\left(1 + \eta_*^{-1} \cdot O(\mathfrak{e})\right). \tag{74}$$

For the remaining bounds, let us first show

$$\mathbb{E}\frac{\eta_*^2}{(\mathsf{D}^2 + \eta_* - \gamma_*)^2} = 1 + \kappa_2\eta_*^{-2}(1 + \eta_*^{-1} \cdot O(\mathfrak{e})). \tag{75}$$

Note that $|\mathsf{D}^2 - d_*| \le \mathfrak{e}$, $\kappa_2 \le \mathfrak{e}$, and (74) together imply $|\mathsf{D}^2 - \gamma_*| = O(\mathfrak{e})$. Then for all $\mathfrak{e} \le \mathfrak{e}_0(\mathfrak{C})$ sufficiently small,

$$\begin{aligned} \frac{\eta_*^2}{(\mathsf{D}^2 + \eta_* - \gamma_*)^2} &= \left(\frac{1}{1 - \frac{\gamma_* - \mathsf{D}^2}{\eta_*}}\right)^2 \\ &= 1 + \sum_{k=1}^\infty (k+1)\left(\frac{\gamma_* - \mathsf{D}^2}{\eta_*}\right)^k \end{aligned} \tag{76}$$

which is an absolutely convergent series. Let $\mu_j = \mathbb{E}[(d_* - \mathsf{D}^2)^j]$ be the j^{th} central moment of $-\mathsf{D}^2$ (where $\mu_0 = 1$ and $\mu_1 = 0$), which has the bound $|\mu_j| \le \mathfrak{e}^{j-2}\kappa_2$ for all $j \ge 2$ by Lemma 13(c). Applying this bound and $|\gamma_* - d_*| = O(\kappa_2) = O(\mathfrak{e})$, for all $k \ge 3$ and a constant $C = C(\mathfrak{C}) > 0$ we have

$$\begin{aligned} \left|\mathbb{E}\left(\frac{\gamma_* - \mathsf{D}^2}{\eta_*}\right)^k\right| &= \left|\sum_{j=0}^k \binom{k}{j}\eta_*^{-k}(\gamma_* - d_*)^{k-j}\mu_j\right| \\ &\le 2^k\eta_*^{-k}(C\mathfrak{e})^{k-2}\kappa_2. \end{aligned}$$

Then the summation of the terms of (76) for $k \geq 3$ is bounded by $\kappa_2\eta_*^{-3} \cdot O(\mathfrak{e})$ for all $\mathfrak{e} \leq \mathfrak{e}_0(\mathfrak{C})$ sufficiently small. For $k = 1$ and $k = 2$, we have

$$\mathbb{E}\left(\frac{\gamma_* - \mathsf{D}^2}{\eta_*}\right) = \frac{\gamma_* - d_*}{\eta_*} = -\kappa_2\eta_*^{-2} + \kappa_2\eta_*^{-3} \cdot O(\mathfrak{e})$$
$$\mathbb{E}\left(\frac{\gamma_* - \mathsf{D}^2}{\eta_*}\right)^2 = \frac{(\gamma_* - d_*)^2 + \kappa_2}{\eta_*^2} = \kappa_2\eta_*^{-2} + \kappa_2\eta_*^{-4} \cdot O(\mathfrak{e})$$

Applying these to (76) gives (75) as claimed.

Now applying (75) to the definition of κ_* in (25),

$$\kappa_* = \left(\frac{\eta_* - \gamma_*}{\gamma_*}\right)^2 \frac{\kappa_2}{\eta_*^2}\left(1 + \eta_*^{-1} \cdot O(\mathfrak{e})\right)$$
$$= \left(\frac{\eta_* - \gamma_*}{\eta_*}\right)^2 \frac{\kappa_2}{d_*^2}\left(1 + \eta_*^{-1} \cdot O(\mathfrak{e})\right), \tag{77}$$

the second equality applying $\gamma_* = d_*(1+O(\mathfrak{e}))$ as implied by (74) and $\kappa_2 = O(d_*\mathfrak{e})$. Applying the first equality of (77) to the definition of α_*^A in (31), and then applying this and (74) to the definition of α_*^B,

$$\alpha_*^A = \frac{\eta_*}{\sqrt{\kappa_2}}(1 + \eta_*^{-1} \cdot O(\mathfrak{e})),$$
$$\alpha_*^B = \frac{\eta_*^2}{\kappa_2} \cdot (\gamma_* - d_*) \cdot (1 + \eta_*^{-1}O(\mathfrak{e})) = -\eta_* + O(\mathfrak{e}).$$

Inverting (74) and applying $\kappa_2 = d_*\mathfrak{e}$, we have $\gamma_*^{-1} = d_*^{-1}(1+\eta_*^{-1}O(\kappa_2/d_*))$. Then applying (77) and $(\eta_* - \gamma_*)/\eta_* \in (0,1)$ where this lower bound of 0 follows from Proposition 1, we obtain from the definition of b_* in (25) that

$$b_* = \frac{1}{\gamma_*} - \frac{\kappa_*}{\eta_* - \gamma_*}$$
$$= \frac{1}{d_*}\left(1 + \eta_*^{-1} \cdot O\left(\frac{\kappa_2}{d_*}\right)\right) + \frac{1}{d_*\eta_*} \cdot O\left(\frac{\kappa_2}{d_*}\right)$$
$$= \frac{1}{d_*}\left(1 + \eta_*^{-1} \cdot O\left(\frac{\kappa_2}{d_*}\right)\right).$$

Applying $\kappa_2/d_* = O(\mathfrak{e})$, we have from (77) that $\kappa_* = d_*^{-1}O(\mathfrak{e})$. Then

$$\frac{e_*}{b_*} = \frac{1+\kappa_*}{b_*} = d_*(1 + O(\kappa_2/d_*))(1 + d_*^{-1}O(\mathfrak{e})) = d_* + O(\mathfrak{e}).$$

Applying $d_*\gamma_*^{-1} = 1 + \eta_*^{-1}O(\kappa_2/d_*)$ and (77), we have from the definition of a_* in (31) that

$$\frac{a_*}{\sqrt{\kappa_*}} = \frac{(\eta_* - \gamma_*)(1 - d_*\gamma_*^{-1})}{\sqrt{\kappa_*}}$$
$$= O\left(\frac{(\eta_* - \gamma_*)/(d_*\eta_*) \cdot \kappa_2}{\sqrt{\kappa_*}}\right)$$
$$= O(\sqrt{\kappa_2}) = O(\mathfrak{e}).$$

Applying $d_*\gamma_*^{-1} = 1 + \eta_*^{-1}O(\kappa_2/d_*) = 1 + O(\mathfrak{e})$, $d_* - \gamma_* = \kappa_2\eta_*^{-1}(1 + \eta_*^{-1}O(\mathfrak{e}))$, and (77), we have from the definition of c_* in (31) that

$$\frac{c_*}{\kappa_*} = -(\eta_* - \gamma_*) + \frac{1}{\kappa_*}\left(\frac{\eta_* - \gamma_*}{\gamma_*}\right)^2 (d_* - \gamma_*)$$
$$= -(\eta_* - \gamma_*) + \frac{\eta_*^2(d_* - \gamma_*)}{\kappa_2}(1 + \eta_*^{-1} \cdot O(\mathfrak{e}))$$
$$= -(\eta_* - \gamma_*) + \eta_*(1 + \eta_*^{-1} \cdot O(\mathfrak{e})) = d_* + O(\mathfrak{e}).$$

■

D. Convergence of Stationary VAMP

Lemma 8: Recall the replica-symmetric potential $i_{\mathrm{RS}}(\eta^{-1}, \gamma)$ from (12), and let $(\eta_*^{-1}, \gamma_*) \in (0, G(-d_-)) \times \mathbb{R}_+$ be any fixed point of (11) for which γ_* is a local minimizer of

$$\gamma \mapsto \sup_{\eta^{-1} \in (0, G(-d_-))} i_{\mathrm{RS}}(\eta^{-1}, \gamma)$$

Then the conclusion $\lim_{\min(s,t)\to\infty} \delta_{st} = \delta_*$ of Proposition 4 holds for the stationary initialization (21).

Proof: Recall that $\delta_{tt} = \delta_*$ for all $t \geq 1$ from Theorem 2.2. Then $\delta_{st} = \mathbb{E}[\mathsf{X}_s\mathsf{X}_t] \leq \sqrt{\mathbb{E}[\mathsf{X}_s^2]\,\mathbb{E}[\mathsf{X}_t^2]} = \delta_*$ for all $s, t \geq 1$. For $s = 1$ and any $t \geq 2$, observe also that

$$\delta_{1t} = \mathbb{E}\mathsf{X}_1\mathsf{X}_t = \mathbb{E}\left[F(\mathsf{P}_0, \mathsf{X}^\star)\,F(\mathsf{Y}_{t-1} + \mathsf{E}, \mathsf{X}^\star)\right]$$
$$= \mathbb{E}\left[\mathbb{E}\left[F(\mathsf{P}_0, \mathsf{X}^\star)\,F(\mathsf{Y}_{t-1} + \mathsf{E}, \mathsf{X}^\star) \mid \mathsf{X}^\star\right]\right]$$
$$= \mathbb{E}[\mathbb{E}\left[F(\mathsf{P}_0, \mathsf{X}^\star) \mid \mathsf{X}^\star\right]^2] \geq 0 \tag{78}$$

where the last equality holds because P_0, $\mathsf{Y}_{t-1} + \mathsf{E}$, and $\mathsf{X}^\star$ are independent, with P_0 and $\mathsf{Y}_{t-1} + \mathsf{E}$ equal in law (by the identity $\sigma_*^2 + b_* = \gamma_*^{-1}$).

Consider now the map $\delta_{st} \mapsto \delta_{s+1,t+1}$. Recalling that $\mathbb{E}\mathsf{Y}_t^2 = \sigma_*^2$ and $\mathbb{E}\mathsf{Y}_s\mathsf{Y}_t = \kappa_*\delta_{st}$, we may represent

$$(\mathsf{Y}_s + \mathsf{E}, \mathsf{Y}_t + \mathsf{E}) \overset{L}{=} \Big(\sqrt{\kappa_*\delta_{st} + b_*}\mathsf{G} + \sqrt{\sigma_*^2 - \kappa_*\delta_{st}}\mathsf{G}',$$
$$\sqrt{\kappa_*\delta_{st} + b_*}\mathsf{G} + \sqrt{\sigma_*^2 - \kappa_*\delta_{st}}\mathsf{G}''\Big)$$

where $\mathsf{G}, \mathsf{G}', \mathsf{G}''$ are jointly independent standard Gaussian variables. Denote

$$\mathsf{P}_\delta' := \sqrt{\kappa_*\delta + b_*} \cdot \mathsf{G} + \sqrt{\sigma_*^2 - \kappa_*\delta} \cdot \mathsf{G}',$$
$$\mathsf{P}_\delta'' := \sqrt{\kappa_*\delta + b_*} \cdot \mathsf{G} + \sqrt{\sigma_*^2 - \kappa_*\delta} \cdot \mathsf{G}''$$

and define $g : [0, \delta_*] \to \mathbb{R}$ by $g(\delta) := \mathbb{E}\left[F(P_\delta', \mathsf{X}^\star)\,F(P_\delta'', \mathsf{X}^\star)\right]$. Then $\delta_{s+1,t+1} = g(\delta_{st})$.

We claim that for any $\delta \in [0, \delta_*]$, we have $g(\delta) \geq 0$, $g'(\delta) \geq 0$, and $g''(\delta) \geq 0$. The first bound $g(\delta) \geq 0$ follows from

$$g(\delta) = \mathbb{E}\Big[\mathbb{E}[F(\mathsf{P}_\delta', \mathsf{X}^\star)\,F(\mathsf{P}_\delta'', \mathsf{X}^\star) \mid \mathsf{X}^\star, \mathsf{G}]\Big]$$
$$= \mathbb{E}\left[\mathbb{E}\left[F(\mathsf{P}_\delta', \mathsf{X}^\star) \mid \mathsf{X}^\star, \mathsf{G}\right]^2\right] \geq 0,$$

because $\mathsf{P}_\delta', \mathsf{P}_\delta''$ are independent and equal in law conditional on $\mathsf{G}, \mathsf{X}^\star$. Differentiating in δ and applying Gaussian integration by parts,

$$g'(\delta) = 2\mathbb{E}\Bigg[F'(\mathsf{P}_\delta', \mathsf{X}^\star)\,F(\mathsf{P}_\delta'', \mathsf{X}^\star)$$
$$\times\left(\frac{\kappa_*}{2\sqrt{\kappa_*\delta + b_*}} \cdot \mathsf{G} - \frac{\kappa_*}{2\sqrt{\sigma_*^2 - \kappa_*\delta}} \cdot \mathsf{G}'\right)\Bigg]$$
$$= \frac{\kappa_*}{\sqrt{\kappa_*\delta + b_*}}\mathbb{E}\left[F'(\mathsf{P}_\delta', \mathsf{X}^\star)\,F(\mathsf{P}_\delta'', \mathsf{X}^\star)\,\mathsf{G}\right]$$
$$- \frac{\kappa_*}{\sqrt{\sigma_*^2 - \kappa_*\delta}}\mathbb{E}\left[F'(\mathsf{P}_\delta', \mathsf{X}^\star)\,F(\mathsf{P}_\delta'', \mathsf{X}^\star)\,\mathsf{G}'\right]$$
$$= \kappa_*\mathbb{E}\Big[F''(\mathsf{P}_\delta', \mathsf{X}^\star)\,F(\mathsf{P}_\delta'', \mathsf{X}^\star) + F'(\mathsf{P}_\delta', \mathsf{X}^\star)\,F'(\mathsf{P}_\delta'', \mathsf{X}^\star)\Big]$$
$$- \kappa_*\mathbb{E}\left[F''(\mathsf{P}_\delta', \mathsf{X}^\star)\,F(\mathsf{P}_\delta'', \mathsf{X}^\star)\right]$$
$$= \kappa_*\mathbb{E}\left[F'(\mathsf{P}_\delta', \mathsf{X}^\star)\,F'(\mathsf{P}_\delta'', \mathsf{X}^\star)\right].$$

Then $g'(\delta) = \kappa_* \mathbb{E}\left[\mathbb{E}[F'(\mathsf{P}'_\delta, \mathsf{X}^\star) \mid \mathsf{G}, \mathsf{X}^\star]^2\right] \geq 0$, and a similar argument shows $g''(\delta) \geq 0$.

Observe that at $\delta = \delta_*$, we have $\mathsf{P}'_{\delta_*} = \mathsf{P}''_{\delta_*} = \sqrt{\sigma_*^2 + b_*} \cdot \mathsf{G} = \mathsf{G}/\sqrt{\gamma_*}$ which is equal in law to $\mathsf{P} \sim N(0, \gamma_*^{-1})$. Then $g(\delta_*) = \mathbb{E}[F(\mathsf{P}, \mathsf{X}^\star)^2] = \delta_*$ by (62). So $g : [0, \delta_*] \to [0, \delta_*]$ is a nonnegative, increasing, convex function with a fixed point at δ_*. We claim that

$$g'(\delta_*) < 1 \tag{79}$$

This then implies that δ_* is the unique fixed point of $g(\cdot)$ over $[0, \delta_*]$, and $\lim_{t\to\infty} g^{(t)}(\delta) = \delta_*$ for any $\delta \in [0, \delta_*]$. Observe from (78) that $\delta_{1t} = \delta_{12}$ for all $t \geq 2$, so $\delta_{t,t+s} = g^{(t-1)}(\delta_{1,1+s}) = g^{(t-1)}(\delta_{12})$ for any $s \geq 1$. Then $\lim_{\min(s,t)\to\infty} \delta_{st} = \delta_*$ follows.

It remains to show (79). For this, applying (61) and $\mathbb{E}[F'(\mathsf{P}, \mathsf{X}^\star)] = 0$ by (62), we have

$$\begin{aligned} g'(\delta_*) &= \kappa_* \mathbb{E}[F'(\mathsf{P}'_{\delta_*}, \mathsf{X}^\star)^2] \\ &= \kappa_* \left(\frac{\eta_*^2}{(\eta_* - \gamma_*)^2} \mathbb{E}[f'(\mathsf{P} + \mathsf{X}^\star, \gamma_*)^2] - \frac{\gamma_*^2}{(\eta_* - \gamma_*)^2} \right). \end{aligned} \tag{80}$$

Recall the replica-symmetric potential i_{RS} from (12), whose gradient and Hessian in (γ, η^{-1}) are given by

$$\nabla i_{\mathrm{RS}}(\eta^{-1}, \gamma) = \frac{1}{2}\Big(\mathrm{mmse}(\gamma) - \eta^{-1} \quad - R(\eta^{-1}) - \gamma\Big)$$
$$\nabla^2 i_{\mathrm{RS}} = \frac{1}{2}\begin{pmatrix} \mathrm{mmse}'(\gamma) & -1 \\ -1 & -R'(\eta^{-1}) \end{pmatrix}$$

where we have used the I-MMSE relation $i'(\gamma) = \frac{1}{2}\mathrm{mmse}(\gamma)$ [38]. The condition that (η_*^{-1}, γ_*) is a fixed point of (11) implies $\nabla i_{\mathrm{RS}}(\eta_*^{-1}, \gamma_*) = 0$. Here, $\eta^{-1} \mapsto i_{\mathrm{RS}}(\eta^{-1}, \gamma_*)$ is concave because $-R'(\eta^{-1}) > 0$ by Lemma 13(b), so η_*^{-1} is the global maximizer of this function. Then, the given assumption that γ_* is a local minimizer of $\gamma \mapsto \sup_{\eta^{-1}} i_{\mathrm{RS}}(\eta^{-1}, \gamma)$ implies the Schur-complement condition for $\nabla^2 i_{\mathrm{RS}}$

$$\mathrm{mmse}'(\gamma) + \frac{1}{R'(\eta^{-1})} \geq 0. \tag{81}$$

Differentiating $R(z) = G^{-1}(z) - z^{-1}$ where $G(z) = \mathbb{E}[(z + \mathsf{D}^2)^{-1}]$, and applying κ_* from (25),

$$\begin{aligned} R'(\eta_*^{-1}) &= \frac{1}{G'(G^{-1}(\eta_*^{-1}))} + \eta_*^2 \\ &= \frac{1}{G'(\eta_* - \gamma_*)} + \eta_*^2 \\ &= \eta_*^2 - \frac{1}{\mathbb{E}[(\mathsf{D}^2 + \eta_* - \gamma_*)^{-2}]} \\ &= \eta_*^2 - \frac{\eta_*^2}{(\frac{\eta_*}{\eta_* - \gamma_*})^2 \kappa_* + 1}. \end{aligned}$$

Then

$$\frac{1}{R'(\eta_*^{-1})} = \frac{1}{\eta_*^2}\left(1 + \frac{(\eta_* - \gamma_*)^2}{\eta_*^2 \kappa_*}\right).$$

Recalling also $\mathrm{mmse}'(\gamma) = -\mathbb{E}[\mathbb{V}[\mathsf{X}^\star \mid \mathsf{Y}]^2] = -\gamma_*^{-2}\mathbb{E}[f'(\mathsf{P} + \mathsf{X}^\star, \gamma_*)^2]$ from (70) and Proposition 14, the condition (81) may be rearranged as

$$\mathbb{E}[f'(\mathsf{P} + \mathsf{X}^\star, \gamma_*)^2] \leq \frac{\gamma_*^2}{\eta_*^2}\left(1 + \frac{(\eta_* - \gamma_*)^2}{\eta_*^2 \kappa_*}\right).$$

Substituting into (80) gives $g'(\delta_*) \leq \gamma_*^2/\eta_*^2 < 1$ where the second inequality applies $\eta_* - \gamma_* > 0$ from Proposition 1. This shows the desired claim (79), concluding the proof. ∎

Proof of Proposition 4: For $\mathfrak{e} < \mathfrak{e}_0$ sufficiently small, we have shown in the proof of Proposition 1 that $h'(\eta_*^{-1}) < 1$ strictly where $h'(\eta^{-1}) = -\mathrm{mmse}'(\gamma) \cdot R'(\eta^{-1})$ is as defined in (71). Rearranging this gives $\mathrm{mmse}'(\gamma_*) + 1/R'(\eta_*^{-1}) > 0$, i.e. γ_* is a (strict) local minimizer of $\gamma \mapsto \sup_{\eta^{-1}} i_{\mathrm{RS}}(\eta^{-1}, \gamma)$, so the result follows from Lemma 8. ∎

Appendix C
Analysis of the Conditional First Moment

In this appendix we prove Lemmas 1 and 2. The arguments are extensions of those of [33, Lemmas 3.1 and 3.2], and we will refer to [33] for some of the technical details.

Lemma 9: Let π be any probability distribution over $\mathbb{R}$. Let $c_\pi(a, b)$ be as defined in (29), and let

$$\bar{a} = \sup\left\{a \in \mathbb{R} : \int e^{ax^2} d\pi(x) < \infty\right\} \in [0, \infty].$$

Then for all $a > \bar{a}$ and $b \in \mathbb{R}$, we have $c_\pi(a, b) = \infty$. For each fixed $a < \bar{a}$, the function $b \mapsto \log c_\pi(a, b)$ is continuous and satisfies, for some (a, π)-dependent constant $C > 0$ and for all $b \in \mathbb{R}$,

$$\log c_\pi(a, b) \leq C(b^2 + 1).$$

Proof: For the first statement, suppose $\bar{a} < \infty$, and consider $a > \bar{a}$. Taking $\varepsilon > 0$ such that $a - \varepsilon > \bar{a}$, we have

$$\begin{aligned} c_\pi(a, b) &\geq \int_{|x| \geq |b|/\varepsilon} e^{ax^2 + bx} d\pi(x) \\ &\geq \int_{|x| \geq |b|/\varepsilon} e^{(a-\varepsilon)x^2} d\pi(x) = \infty. \end{aligned}$$

For the second statement, consider any $\bar{a} \in [0, \infty]$ and fix $a < \bar{a}$. Taking $\varepsilon > 0$ such that $a + \varepsilon < \bar{a}$, we have

$$\begin{aligned} c_\pi(a, b) &= \int_{|x| < b/\varepsilon} e^{ax^2 + bx} d\pi(x) + \int_{|x| \geq b/\varepsilon} e^{ax^2 + bx} d\pi(x) \\ &\leq e^{b^2/\varepsilon} \int e^{ax^2} d\pi(x) + \int e^{(a+\varepsilon)x^2} d\pi(x) \\ &\leq C(e^{b^2/\varepsilon} + 1) \leq 2Ce^{b^2/\varepsilon} \end{aligned}$$

for a constant $C = C(a, \varepsilon, \pi) > 0$. Then $\log c_\pi(a, b) \leq C'(b^2 + 1)$ for a constant $C' = C'(a, \varepsilon, \pi) > 0$. In particular, $\log c_\pi(a, b) < \infty$, and continuity in b follows from a standard application of the dominated convergence theorem. ∎

Proof: [Proof of Lemma 1] Recall the $n \times t$ matrices $X_t = (x^1, \ldots, x^t)$, $Y_t = (y^1, \ldots, y^t)$, and $S_t = (s^1, \ldots, s^t)$ which collect the AMP iterates. We fix t and write $\mathcal{G}, X, Y, S, \Delta$ for $\mathcal{G}_t, X_t, Y_t, S_t, \Delta_t$.

From the definition of $\mathcal{Z}(\mathcal{U})$ in (27), applying $y = A\beta^\star + \epsilon = Q^\top DO\beta^\star + \epsilon$ and $\xi = Q\epsilon$, and writing as shorthand $\tilde{\sigma} := \tilde{\sigma}(\sigma) = \sigma - \beta^\star$, we have

$$\mathbb{E}[\mathcal{Z}(\mathcal{U}) \mid \mathcal{G}] \tag{82a}$$
$$= \int \mathbb{I}\Big(\frac{1}{n}\|\tilde{\sigma}\|^2 \in \mathcal{U}\Big) \cdot \exp\left(-\frac{\|\epsilon\|^2}{2} + \frac{n}{2} \cdot f_n(\tilde{\sigma})\right) \prod_{i=1}^n d\pi(\sigma_i) \tag{82b}$$

with

$$f_n(\tilde\sigma) := \frac{2}{n}\log\mathbb{E}\left[\exp\left(-\frac{\tilde\sigma^\top O^\top D^\top DO\tilde\sigma}{2}+\tilde\sigma^\top O^\top D^\top\xi\right)\;\middle|\;\mathcal{G}\right].$$

Conditional on $\mathcal{G}$, the only random quantity in the expectation in (C) is the matrix O. By definition of $\mathcal{G}$ in (28) and the AMP iterations in (24), its conditional law is that of a $\mathrm{Haar}(\mathbb{SO}(n))$ matrix O conditioned on the event $(e_b, S, \Lambda S) = O(e, X, Y)$. Then by [34, Lemma B.2], we may represent this conditional law of O as

$$O|_{\mathcal{G}} \overset{L}{=} (e_b, S, \Lambda S)\begin{pmatrix} e^\top e & e^\top X & e^\top Y\\ X^\top e & X^\top X & X^\top Y\\ Y^\top e & Y^\top X & Y^\top Y\end{pmatrix}^{-1}(e,X,Y)^\top + \Pi_{(e_b,S,\Lambda S)^\perp}\tilde O \Pi^\top_{(e,X,Y)^\perp} \tag{83}$$

where $\Pi_{(e,X,Y)^\perp}, \Pi_{(e_b,S,\Lambda S)^\perp} \in \mathbb{R}^{n\times(n-2t-1)}$ have orthonormal columns orthogonal to the column spans of $(e,X,Y), (e_b,S,\Lambda S) \in \mathbb{R}^{n\times(2t+1)}$ respectively, and $\tilde O \sim \mathrm{Haar}(\mathbb{SO}(n-2t-1))$ is independent of $\mathcal{G}$. We remark that the matrix inverse in (83) is well-defined almost surely for all large n, by Corollary 1 and the statements $\Delta \succ 0$ in Theorem 2.2 and $b_*, \kappa_* > 0$ strictly in Remark 1.1. Writing as shorthand $\Pi = \Pi_{(e_b,S,\Lambda S)^\perp}$ and

$$\tilde\sigma_\perp = \Pi^\top_{(e,X,Y)^\perp}\tilde\sigma \in \mathbb{R}^{n-2t-1},$$
$$\tilde\sigma_\parallel = (e_b,S,\Lambda S)\begin{pmatrix} e^\top e & e^\top X & e^\top Y\\ X^\top e & X^\top X & X^\top Y\\ Y^\top e & Y^\top X & Y^\top Y\end{pmatrix}^{-1}(e,X,Y)^\top\tilde\sigma \in \mathbb{R}^n, \tag{84}$$

this yields the equality in conditional law $O\tilde\sigma|_{\mathcal{G}} \overset{L}{=} \Pi\tilde O\tilde\sigma_\perp + \tilde\sigma_\parallel$. So $f_n(\tilde\sigma)$ in (C) is given by

$$f_n(\tilde\sigma) = \frac{2}{n}\log\mathbb{E}\exp\Bigg(-\frac{(\Pi\tilde O\tilde\sigma_\perp + \tilde\sigma_\parallel)^\top D^\top D(\Pi\tilde O\tilde\sigma_\perp + \tilde\sigma_\parallel)}{2} \tag{85}$$
$$+\left(\Pi\tilde O\tilde\sigma_\perp + \tilde\sigma_\parallel\right)^\top D^\top\xi\Bigg) \tag{86}$$
$$= -\frac{1}{n}\tilde\sigma_\parallel^\top D^\top D\tilde\sigma_\parallel + \frac{2}{n}\tilde\sigma_\parallel^\top D^\top\xi \tag{87}$$
$$+\frac{2}{n}\log\mathbb{E}\exp\Bigg(-\frac{\tilde\sigma_\perp^\top\tilde O^\top\Pi^\top D^\top D\Pi\tilde O\tilde\sigma_\perp}{2} + \left(\xi^\top D\Pi - \tilde\sigma_\parallel^\top D^\top D\Pi\right)\tilde O\tilde\sigma_\perp\Bigg) \tag{88}$$

where this expectation is over only $\tilde O \sim \mathrm{Haar}(\mathbb{SO}(n-2t-1))$.

Uniform Approximation of $f_n(\tilde\sigma)$: We proceed to approximate $f_n(\tilde\sigma)$ by low-dimensional functions of $\tilde\sigma$ for large n. Define $\mathfrak{P}(\tilde\sigma) = (u(\tilde\sigma), r(\tilde\sigma), v(\tilde\sigma), w(\tilde\sigma))$ where

$$u(\tilde\sigma) = \frac{1}{n}\|\tilde\sigma\|^2,$$
$$\begin{pmatrix} r(\tilde\sigma)\\ v(\tilde\sigma)\\ w(\tilde\sigma)\end{pmatrix} = \left[\frac{1}{n}\begin{pmatrix} e^\top e & e^\top X & e^\top Y\\ X^\top e & X^\top X & X^\top Y\\ Y^\top e & Y^\top X & Y^\top Y\end{pmatrix}\right]^{-1/2} \times\frac{1}{n}(e,X,Y)^\top\tilde\sigma. \tag{89}$$

Here, $u(\tilde\sigma), r(\tilde\sigma) \in \mathbb{R}$ and $v(\tilde\sigma), w(\tilde\sigma) \in \mathbb{R}^t$. Note that $O\tilde\sigma_\parallel$ is the orthogonal projection of $\tilde\sigma$ onto the column span of (e,X,Y), so

$$r^2(\tilde\sigma) + \|v(\tilde\sigma)\|^2 + \|w(\tilde\sigma)\|^2 = \frac{\|\tilde\sigma_\parallel\|^2}{n} = u(\tilde\sigma) - \frac{\|\tilde\sigma_\perp\|^2}{n}. \tag{90}$$

Let $K > 0$ be the bound given in the lemma for which $\mathcal{U} \subseteq (0, K)$, and define the open domain

$$\mathcal{K} = \{\tilde\sigma \in \mathbb{R}^n : u(\tilde\sigma) \in (0,K),\ \mathcal{A}(\mathfrak{P}(\tilde\sigma)) > 0\} \tag{91}$$

where $\mathcal{A}(\mathfrak{P}) = u - r^2 - \|v\|^2 - \|w\|^2$ as defined in (32). Restricting first to $\tilde\sigma \in \mathcal{K}$, we apply [33, Proposition 2.7] to approximate the expectation over $\tilde O$. (This is stated in [33] for $\mathrm{Haar}(\mathbb{O}(n))$, but the result and proof hold verbatim for $\mathbb{SO}(n)$.) The needed conditions of [33, Proposition 2.7] are verified by the Cauchy interlacing of eigenvalues of $\Pi^\top D^\top D\Pi$ with those of $D^\top D$, the convergence assumptions (4), the bounds $\|\tilde\sigma_\perp\|^2, \|\tilde\sigma_\parallel\|^2 \le \|\tilde\sigma\|^2 < nK$ for $\tilde\sigma \in \mathcal{K}$, the bound $\|D\xi\|^2 < 2d_+ n$ almost surely for all large n, and the observation that by Lemma 13(c), for $\mathfrak{e} < \mathfrak{e}_0(\mathfrak{C}, K)$ sufficiently small and for some sufficiently small constant $\mathfrak{d} > 0$,

$$G(-d_- + \mathfrak{d}) - \mathfrak{d} > K. \tag{92}$$

(We allow $\mathfrak{d} > 0$ to depend on $\mathfrak{e}$ and the law of D^2, and we will eventually take $\mathfrak{d} \to 0$.) For scalars or vectors $a_n(\tilde\sigma)$ and $b_n(\tilde\sigma)$ whose dimensions are independent of n, let us write $a_n(\tilde\sigma) \doteq b_n(\tilde\sigma)$ to mean, almost surely,

$$\lim_{n,m\to\infty}\sup_{\tilde\sigma\in\mathcal{K}}|a_n(\tilde\sigma) - b_n(\tilde\sigma)| = 0,$$
$$\lim_{n,m\to\infty}\sup_{\tilde\sigma\in\mathcal{K}}\|a_n(\tilde\sigma) - b_n(\tilde\sigma)\| = 0.$$

Then [33, Proposition 2.7] applied to the expectation over $\tilde O$ yields

$$f_n(\tilde\sigma) \doteq -\frac{1}{n}\tilde\sigma_\parallel^\top D^\top D\tilde\sigma_\parallel + \frac{2}{n}\tilde\sigma_\parallel^\top D^\top\xi + E_n(\tilde\sigma) \tag{93}$$

where

$$E_n(\tilde\sigma) = \inf_{\zeta \ge -d_- + \mathfrak{d}}\Bigg\{\frac{\zeta\|\tilde\sigma_\perp\|^2}{n} + n^{-1}(\xi^\top D\Pi - \tilde\sigma_\parallel^\top D^\top D\Pi) \times(\zeta I + \Pi^\top D^\top D\Pi)^{-1}(\Pi^\top D^\top\xi - \Pi^\top D^\top D\tilde\sigma_\parallel) - \frac{1}{n}\log\det\left(\zeta I + \Pi^\top D^\top D\Pi\right) - \left(1 + \log\frac{\|\tilde\sigma_\perp\|^2}{n}\right)\Bigg\}. \tag{94}$$

For the first term of (93), applying Corollary 1 and the preceding definitions of $\tilde{\sigma}_{\|}$ and r, v, w,

$$\tilde{\sigma}_{\|} = e_b \big(b_*^{-1/2} r(\tilde{\sigma})\big) + S\Delta^{-1/2} v(\tilde{\sigma}) + \Lambda S\,(\kappa_* \Delta)^{-1/2} w(\tilde{\sigma}) + r_n(\tilde{\sigma}) \quad (95)$$

where $n^{-1}\|r_n(\tilde{\sigma})\|^2 \to 0$ uniformly over $\tilde{\sigma} \in \mathcal{K}$. It follows from Theorem 2.2, Corollaries 1 and 2, $\mathbb{E}\mathsf{D}^2 = d_*$, and (65) that almost surely,

$$\lim_{n,m\to\infty} \frac{1}{n}(e_b, S, \Lambda S)^\top D^\top D(e_b, S, \Lambda S) = \begin{pmatrix} e_* & 0 & 0 \\ 0 & d_*\Delta & a_*\Delta \\ 0 & a_*\Delta & c_*\Delta \end{pmatrix}. \quad (96)$$

Combining this with (95), we obtain for the first term of (93) that

$$\frac{1}{n}\tilde{\sigma}_{\|}^\top D^\top D \tilde{\sigma}_{\|} \doteq \frac{e_* r^2(\tilde{\sigma})}{b_*} + \operatorname{Tr}\begin{pmatrix} d_* & a_* \\ a_* & c_* \end{pmatrix} \left(v(\tilde{\sigma}), \frac{w(\tilde{\sigma})}{\sqrt{\kappa_*}}\right)^\top \left(v(\tilde{\sigma}), \frac{w(\tilde{\sigma})}{\sqrt{\kappa_*}}\right). \quad (97)$$

Similarly, for the second term of (93), applying the form of E_b in Proposition 3 and (64b), we have

$$\frac{2}{n}\tilde{\sigma}_{\|}^\top D\xi \doteq \frac{2}{n} e_b^\top D\xi \left(b_*^{-1/2} r(\tilde{\sigma})\right) \doteq \frac{2r(\tilde{\sigma})}{\sqrt{b_*}} \quad (98)$$

where the contributions from the other terms of $\tilde{\sigma}_{\|}$ vanish because Ξ in Proposition 3 has mean 0 and is independent of $(\mathsf{S}_1, \ldots, \mathsf{S}_t)$ and (D, L).

For the final term $E_n(\tilde{\sigma})$ of (93), note that (4) and Cauchy eigenvalue interlacing imply that the empirical eigenvalue distribution of $\Pi^\top D^\top D \Pi$ converges weakly to D^2. Then, recalling $n^{-1}\|\tilde{\sigma}_\perp\|^2 = u(\tilde{\sigma}) - r^2(\tilde{\sigma}) - \|v(\tilde{\sigma})\|^2 - \|w(\tilde{\sigma})\|^2 = \mathcal{A}(\mathfrak{P}(\tilde{\sigma}))$ from (90) and the definition of $\mathcal{H}$ in (33),

$$\frac{\zeta\|\tilde{\sigma}_\perp\|^2}{n} - \frac{1}{n}\log\det\left(\zeta I + \Pi^\top D^\top D\Pi\right) - \left(1 + \log\frac{\|\tilde{\sigma}_\perp\|^2}{n}\right) \doteq \mathcal{H}\left(\zeta, \mathcal{A}(\mathfrak{P}(\tilde{\sigma}))\right). \quad (99)$$

The error of this approximation converges to 0 uniformly over $\zeta \geq -d_- + \mathfrak{d}$, by the same Arzelà-Ascoli argument as leading to [33, Eq. (3.25)].

To analyze the remaining second term of $E_n(\tilde{\sigma})$ in (94), let us define

$$\bar{\Pi} = (e_b, S, \Lambda S)\begin{pmatrix} e_b^\top e_b & e_b^\top S & e_b^\top \Lambda S \\ S^\top e_b & S^\top S & S^\top \Lambda S \\ S^\top \Lambda e_b & S^\top \Lambda S & S^\top \Lambda^2 S \end{pmatrix}^{-1/2} = (e_b, S, \Lambda S)\begin{pmatrix} e^\top e & e^\top X & e^\top Y \\ X^\top e & X^\top X & X^\top Y \\ Y^\top e & Y^\top X & Y^\top Y \end{pmatrix}^{-1/2} \quad (100)$$

whose columns are an orthogonalization of $(e_b, S, \Lambda S)$. Then the columns of $(\Pi, \bar{\Pi})$ form a full orthonormal basis for $\mathbb{R}^n$. Applying the Schur-complement formula for block matrix inversion, we obtain analogously to [33, Eq. (3.29)] that the second term of (94) is given by

$$n^{-1}(\xi^\top D\Pi - \tilde{\sigma}_{\|}^\top D^\top D\Pi)(\zeta I + \Pi^\top D^\top D\Pi)^{-1} \times (\Pi^\top D^\top \xi - \Pi^\top D^\top D\tilde{\sigma}_{\|}) = \mathrm{I} - \mathrm{II}$$

where

$$\mathrm{I} = n^{-1}(\xi^\top D - \tilde{\sigma}_{\|}^\top D^\top D)\Pi \cdot \Pi^\top \left(\zeta I + D^\top D\right)^{-1}\Pi \quad (101)$$

$$\times\, \Pi^\top \left(D^\top \xi - D^\top D\tilde{\sigma}_{\|}\right) \quad (102)$$

$$\mathrm{II} = n^{-1}(\xi^\top D - \tilde{\sigma}_{\|}^\top D^\top D)\Pi \quad (103)$$

$$\times\, \Pi^\top(\zeta I + D^\top D)^{-1}\bar{\Pi}\left(\bar{\Pi}^\top\left(\zeta I + D^\top D\right)^{-1}\bar{\Pi}\right)^{-1} \times \bar{\Pi}^\top(\zeta I + D^\top D)^{-1}\Pi \cdot \Pi^\top(D^\top \xi - D^\top D\tilde{\sigma}_{\|}) \quad (104)$$

We derive almost-sure asymptotic limits for I and II. Recalling $\lambda(\cdot)$ and $\theta(\cdot)$ from (33), let us define

$$\mathcal{F}^e(\zeta, r) = \mathcal{F}^e_{22}(\zeta, r) - \mathcal{F}^e_{12}(\zeta, r)\mathcal{F}^e_{11}(\zeta)^{-1}\mathcal{F}^e_{12}(\zeta, r)$$
$$\mathcal{F}(\zeta) = \mathcal{F}_{22}(\zeta) - \mathcal{F}_{12}(\zeta)^\top \mathcal{F}_{11}(\zeta)^{-1}\mathcal{F}_{12}(\zeta) \quad (105)$$

where we set

$$\upsilon^A(x) = b_*^{-1/2}\left(-x + \frac{e_*}{b_*}\right), \quad (106)$$

$$\upsilon^B(x) = \left(\frac{\gamma_*}{\eta_*}\right)x + \frac{\gamma_*(\eta_* - \gamma_*)}{\eta_*} - \frac{1}{b_*}, \quad (107)$$

$$\mathfrak{f}^e(x, r) = \upsilon^A(x)\cdot r + \upsilon^B(x),$$

$$\mathcal{F}^e_{11}(\zeta) = \mathbb{E}\left(\frac{1}{\zeta + \mathsf{D}^2}\mathsf{E}_b^2\right), \quad (108)$$

$$\mathcal{F}^e_{12}(\zeta, r) = \mathbb{E}\left(\frac{\mathfrak{f}^e(\mathsf{D}^2, r)}{\zeta + \mathsf{D}^2}\mathsf{E}_b^2\right), \quad (109)$$

$$\mathcal{F}^e_{22}(\zeta, r) = \mathbb{E}\left(\frac{\mathfrak{f}^e(\mathsf{D}^2, r)^2}{\zeta + \mathsf{D}^2}\mathsf{E}_b^2\right), \quad (110)$$

$$\mathcal{F}_{11}(\zeta) = \mathbb{E}\frac{1}{\zeta + \mathsf{D}^2}\begin{pmatrix} 1 & \lambda(\mathsf{D}^2) \\ \lambda(\mathsf{D}^2) & \lambda(\mathsf{D}^2)^2 \end{pmatrix},$$

$$\mathcal{F}_{12}(\zeta) = \mathbb{E}\frac{1}{\zeta + \mathsf{D}^2}\begin{pmatrix} \theta(\mathsf{D}^2) \\ \lambda(\mathsf{D}^2)\theta(\mathsf{D}^2) \end{pmatrix}, \quad (111)$$

$$\mathcal{F}_{22}(\zeta) = \mathbb{E}\frac{1}{\zeta + \mathsf{D}^2}\theta(\mathsf{D}^2)^2. \quad (112)$$

Since D^2 has strictly positive variance and $x \mapsto \lambda(x)$ is one-to-one on the support of D^2, we have $\mathcal{F}_{11}(\zeta) \succ 0$ strictly for $\zeta > -d_-$, so $\mathcal{F}(\zeta)$ is well-defined. Also $\mathcal{F}^e_{11}(\zeta) > 0$ strictly since $\mathbb{E}\mathsf{E}_b^2 = b_* > 0$ from Theorem 1.1, so $\mathcal{F}^e(\zeta, r)$ is well-defined. We note that these functions are expressed equivalently as

$$\mathcal{F}^e(\zeta, r) = \inf_{\chi^A \in \mathbb{R}} \mathbb{E}\left[\frac{\mathsf{E}_\mathrm{b}^2}{\zeta + \mathsf{D}^2}\left(\mathfrak{f}^e(\mathsf{D}^2, r) - \chi^A\right)^2\right] = \inf_{\chi^A \in \mathbb{R}} \mathbb{E}\left[\frac{\mathsf{E}_\mathrm{b}^2}{\zeta + \mathsf{D}^2}\left(\left[\frac{\gamma_*}{\eta_*} - b_*^{-1/2} r\right]\mathsf{D}^2 - \chi^A\right)^2\right]$$
$$\mathcal{F}(\zeta) = \inf_{\chi^B, \chi^C \in \mathbb{R}} \mathbb{E}\left[\frac{1}{\zeta + \mathsf{D}^2}\left(\theta(\mathsf{D}^2) - \lambda(\mathsf{D}^2)\chi^B - \chi^C\right)^2\right] \quad (113)$$

where these coincide with the above definitions upon explicitly evaluating the infima over $\chi^A, \chi^B, \chi^C \in \mathbb{R}$, and the two expressions for $\mathcal{F}^e$ in (113) are identical upon absorbing all terms of $\mathfrak{f}^e(\mathsf{D}^2, r)$ not depending on D^2 into an additive shift of the variable χ^A.

We first approximate the common term $\Pi\Pi^\top(D^\top\xi - D^\top D\tilde{\sigma}_\|)$ in (102–104): Applying Corollary 1 together with (95) and (96), we obtain

$$\begin{pmatrix} e^\top e & e^\top X & e^\top Y \\ X^\top e & X^\top X & X^\top Y \\ Y^\top e & Y^\top X & Y^\top Y \end{pmatrix}^{-1} (e_b, S, \Lambda S)^\top D^\top D\tilde{\sigma}_\|$$
$$\doteq \begin{pmatrix} b_* & 0 & 0 \\ 0 & \Delta & 0 \\ 0 & 0 & \kappa_*\Delta \end{pmatrix}^{-1} \begin{pmatrix} e_* & 0 & 0 \\ 0 & d_*\Delta & a_*\Delta \\ 0 & a_*\Delta & c_*\Delta \end{pmatrix} \begin{pmatrix} b_*^{-1/2} r(\tilde{\sigma}) \\ \Delta^{-1/2} v(\tilde{\sigma}) \\ (\kappa_*\Delta)^{-1/2} w(\tilde{\sigma}) \end{pmatrix}.$$

Then applying

$$\Pi\Pi^\top = I - \bar{\Pi}\bar{\Pi}^\top = I - (e_b, S, \Lambda S) \begin{pmatrix} e^\top e & e^\top X & e^\top Y \\ X^\top e & X^\top X & X^\top Y \\ Y^\top e & Y^\top X & Y^\top Y \end{pmatrix}^{-1} (e_b, S, \Lambda S)^\top \quad (114)$$

yields

$$\Pi\Pi^\top D^\top D\tilde{\sigma}_\| = b_*^{-1/2} r(\tilde{\sigma}) \left(D^\top D - e_* b_*^{-1} I\right) e_b + \left(D^\top D - d_* I - a_*\kappa_*^{-1}\Lambda\right) S\Delta^{-1/2} v(\tilde{\sigma}) + \kappa_*^{-1/2} \left(D^\top D\Lambda - a_* I - c_*\kappa_*^{-1}\Lambda\right) S\Delta^{-1/2} w(\tilde{\sigma}) + r_n(\tilde{\sigma}) \quad (115)$$

where $n^{-1}\|r_n(\tilde{\sigma})\|^2 \to 0$ uniformly over $\tilde{\sigma} \in \mathcal{K}$. From the definitions of a_*, c_* in (31) and of Λ in (22), a straightforward computation yields the identity

$$-\frac{\gamma_*}{\eta_* - \gamma_*}\left(D^\top D\Lambda - a_* I - \frac{c_*}{\kappa_*}\Lambda\right) = D^\top D - d_* I - \frac{a_*}{\kappa_*}\Lambda =: \widetilde{D} \quad (116)$$

where we define $\widetilde{D} \in \mathbb{R}^{n\times n}$ as this common quantity. Then, recalling α_*^A from (31), we can rewrite (115) as

$$\Pi\Pi^\top D^\top D\tilde{\sigma}_\| = b_*^{-1/2} r(\tilde{\sigma}) \cdot \left(D^\top D - e_* b_*^{-1} I\right) e_b + \widetilde{D} S\Delta^{-1/2}\left(v(\tilde{\sigma}) - \alpha_*^A w(\tilde{\sigma})\right) + r_n(\tilde{\sigma}). \quad (117)$$

Applying Corollary 1, (96), (114) and $D^\top\xi = \frac{\gamma_*}{\eta_*}\left[D^\top D + (\eta_* - \gamma_*) I\right] e_b$, we also have similarly to (98)

$$\Pi\Pi^\top D^\top\xi = D^\top\xi - (e_b, S, \Lambda S) \begin{pmatrix} b_* & 0 & 0 \\ 0 & \Delta & 0 \\ 0 & 0 & \kappa_*\Delta \end{pmatrix}^{-1} \times \frac{1}{n}(e_b, S, \Lambda S)^\top D^\top\xi + r_n(\tilde{\sigma})$$
$$= \frac{\gamma_*}{\eta_*}\left[D^\top D + (\eta_* - \gamma_*) I\right] e_b - b_*^{-1} e_b + r_n(\tilde{\sigma}). \quad (118)$$

Then combining (117) and (118), and applying $\mathfrak{f}^e$ from (110) to $D^\top D$ by functional calculus,

$$\Pi\Pi^\top(D^\top\xi - D^\top D\tilde{\sigma}_\|) = \mathfrak{f}^e(D^\top D, r(\tilde{\sigma})) e_b - \widetilde{D} S\Delta^{-1/2}\left(v(\tilde{\sigma}) - \alpha_*^A w(\tilde{\sigma})\right) + r_n(\tilde{\sigma}) \quad (119)$$

for a remainder $r_n(\tilde{\sigma})$ satisfying $n^{-1}\|r_n(\tilde{\sigma})\|^2 \to 0$ uniformly over $\tilde{\sigma} \in \mathcal{K}$.

We now apply (119) and Corollary 2 to approximate the two terms (102) and (104): Recalling α_*^B from (31), observe that the second definition for $\widetilde{D}$ in (116) has the equivalent form

$$\widetilde{D} = D^\top D - \alpha_*^B \eta_* \left(D^\top D + (\eta_* - \gamma_*) I\right)^{-1} + (\alpha_*^B - d_*) I.$$

Then, recalling the definitions of $\mathcal{F}_{22}, \mathcal{F}_{22}^e$ from (110), by Theorem 2.2 and Corollary 2 we have

$$n^{-1} S^\top \widetilde{D}\left(\zeta I + D^\top D\right)^{-1} \widetilde{D} S \doteq \mathcal{F}_{22}(\zeta) \cdot \Delta,$$
$$n^{-1} e_b^\top \mathfrak{f}^e\left(D^\top D, r(\tilde{\sigma})\right)\left(\zeta I + D^\top D\right)^{-1} \mathfrak{f}^e\left(D^\top D, r(\tilde{\sigma})\right) e_b \doteq \mathcal{F}_{22}^e(\zeta, r(\tilde{\sigma})),$$
$$n^{-1} e_b^\top \mathfrak{f}^e((D^\top D, r(\tilde{\sigma}))\left(\zeta I + D^\top D\right)^{-1} \widetilde{D} S \doteq 0.$$

Combining with (119), this shows for (102) that

$$\mathrm{I} \doteq \mathcal{F}_{22}^e(\zeta, r(\tilde{\sigma})) + \mathcal{F}_{22}(\zeta) \cdot \left\|v(\tilde{\sigma}) - \alpha_*^A w(\tilde{\sigma})\right\|_2^2 = \mathcal{F}_{22}^e(\zeta, r(\tilde{\sigma})) + \mathcal{F}_{22}(\zeta) \cdot \mathcal{B}(v(\tilde{\sigma}), w(\tilde{\sigma})). \quad (120)$$

For (104), by Theorem 2.2 and Corollary 2, we have

$$n^{-1}(S, \Lambda S)^\top(\zeta I + D^\top D)^{-1}(S, \Lambda S) \doteq \mathcal{F}_{11}(\zeta) \otimes \Delta \in \mathbb{R}^{2t\times 2\ t},$$
$$n^{-1} e_b^\top(\zeta I + D^\top D)^{-1} e_b \doteq \mathcal{F}_{11}^e(\zeta),$$
$$n^{-1} e_b^\top(\zeta I + D^\top D)^{-1}(S, \Lambda S) \doteq 0.$$

Then, recalling the form of $\bar{\Pi}$ from (100),

$$\bar{\Pi}\left(\bar{\Pi}^\top\left(\zeta I + D^\top D\right)^{-1}\bar{\Pi}\right)^{-1}\bar{\Pi}^\top = n^{-1} \times \left(e_b \mathcal{F}_{11}^e(\zeta)^{-1} e_b^\top + (S, \Lambda S)(\mathcal{F}_{11}(\zeta) \otimes \Delta)^{-1}(S, \Lambda S)^\top + r_n\right)$$

where $n^{-1}\|r_n\| \to 0$ in operator norm. Combining this with (119), and applying Theorem 2.2, Corollary 2, and the definitions of $\mathcal{F}_{12}, \mathcal{F}_{12}^e$ in (110), we obtain for (104)

$$\mathrm{II} \doteq \mathcal{F}_{12}^e(\zeta, r(\tilde{\sigma}))\mathcal{F}_{11}^e(\zeta)^{-1}\mathcal{F}_{12}^e(\zeta, r(\tilde{\sigma})) + \mathcal{F}_{12}(\zeta)^\top \mathcal{F}_{11}(\zeta)^{-1}\mathcal{F}_{12}(\zeta) \cdot \mathcal{B}(v(\tilde{\sigma}), w(\tilde{\sigma})). \quad (121)$$

Combining (120) and (121), the second term of $E_n(\tilde{\sigma})$ in (94) satisfies

$$n^{-1}(\xi^\top D\Pi - \tilde{\sigma}_\|^\top D^\top D\Pi) \times (\zeta I + \Pi^\top D^\top D\Pi)^{-1}(\Pi^\top D^\top\xi - \Pi^\top D^\top D\tilde{\sigma}_\|) \doteq \mathcal{F}^e(\zeta, r(\tilde{\sigma})) + \mathcal{F}(\zeta) \cdot \mathcal{B}(v(\tilde{\sigma}), w(\tilde{\sigma})). \quad (122)$$

The error of this approximation again converges to 0 uniformly over $\zeta \geq -d_- + \mathfrak{d}$, by an argument that is the same as that leading to [33, Eq. (3.31)].

Combining (97), (98), (99), and (122) and applying this to (93), we obtain

$$\lim_{n,m\to\infty} \sup_{\tilde{\sigma}\in\mathcal{K}} \left| f_n(\tilde{\sigma}) - f(\mathfrak{P}(\tilde{\sigma}))\right| = 0,$$

where we define on the domain $\mathcal{V} = \{\mathfrak{P} : u - r^2 - \|v\|^2 - \|w\|^2 > 0\}$ the function

$$f(\mathfrak{P}) = \inf_\zeta -\frac{e_* r^2}{b_*} + \frac{2r}{\sqrt{b_*}} - \mathrm{Tr}\left[\begin{pmatrix} d_* & a_* \\ a_* & c_* \end{pmatrix}\left(v, \frac{w}{\sqrt{\kappa_*}}\right)^\top\left(v, \frac{w}{\sqrt{\kappa_*}}\right)\right] + \mathcal{H}(\zeta, \mathcal{A}(\mathfrak{P})) + \mathcal{F}^e(\zeta, r) + \mathcal{F}(\zeta) \cdot \mathcal{B}(v, w),$$

and the infimum is over $\zeta \geq -d_- + \mathfrak{d}$. The functions $\mathcal{F}^e, \mathcal{F}$ are decreasing over $\zeta > -d_-$ by definition in (113). For any fixed $\mathfrak{P}$ with $u < K$, applying $\frac{\partial}{\partial \zeta}\mathcal{H}(\zeta, \mathcal{A}(\mathfrak{P})) = \mathcal{A}(\mathfrak{P}) - G(\zeta) < K - G(\zeta) < 0$ by (92), the function $\mathcal{H}(\zeta, \mathcal{A}(\mathfrak{P}))$ is also decreasing over $\zeta \in (-d_-, -d_- + \mathfrak{d}]$. Then the above infimum defining $f(\mathfrak{P})$ may be extended to $\zeta > -d_-$. Finally, this uniform approximation for f_n may be extended from $\mathcal{K}$ to its closure $\overline{\mathcal{K}}$: Here f_n as defined in (C) is continuous on $\overline{\mathcal{K}}$, and the map $\mathfrak{P} : \overline{\mathcal{K}} \to \{\mathfrak{P} \in \bar{\mathcal{V}} : u \in [0, K]\}$ is continuous, relatively open, and maps the dense subset $\mathcal{K} \subset \overline{\mathcal{K}}$ to the interior $\{\mathfrak{P} \in \mathcal{V} : u \in (0, K)\}$ for each fixed n. Then [33, Proposition C.1] shows that f has a continuous extension to $\{\mathfrak{P} \in \bar{\mathcal{V}} : u \in [0, K]\}$, and (denoting also by f this extension)

$$\lim_{n,m\to\infty} \sup_{\tilde{\sigma} \in \overline{\mathcal{K}}} |f_n(\tilde{\sigma}) - f(\mathfrak{P}(\tilde{\sigma}))| = 0.$$

Applying this to (82), and denoting by $\langle \cdot \rangle_\pi$ the expectation over $(\sigma_i)_{i=1}^n \overset{iid}{\sim} \pi$, we obtain almost surely

$$\lim_{n,m\to\infty} \frac{1}{n} \log \mathbb{E}[\mathcal{Z}(\mathcal{U}) \mid \mathcal{G}] = -\frac{1}{2} + \lim_{n,m\to\infty} \frac{1}{n} \log \left\langle \mathbb{I}\{u(\tilde{\sigma}) \in \mathcal{U}\} \exp\left(\frac{n}{2} f(\mathfrak{P}(\tilde{\sigma}))\right) \right\rangle_\pi . \tag{123}$$

The same statement also holds with the closure $\overline{\mathcal{U}}$ in place of $\mathcal{U}$ on both sides.

Large Deviations Analysis: We conclude the proof by establishing large deviations upper and lower bounds for $\mathfrak{P}(\tilde{\sigma})$ and applying Varadhan's lemma. Recall that $\tilde{\sigma} = \sigma - \beta^\star$, and introduce dual variables $\mathfrak{R} = (U, R, V, W)$ where $U, R \in \mathbb{R}$ and $V, W \in \mathbb{R}^t$. For the large deviations upper bound, define the cumulant generating function

$$\begin{aligned} \lambda_n(\mathfrak{R}) &= \frac{1}{n} \log \left\langle \exp\left(n \cdot \mathfrak{P}(\tilde{\sigma})^\top \mathfrak{R}\right) \right\rangle_\pi \\ &= \frac{1}{n} \sum_{i=1}^n \log \int \exp\left(U\sigma_i^2 + A_i \sigma_i + B_i\right) d\pi(\sigma_i) \\ &= \frac{1}{n} \sum_{i=1}^n \left[B_i + \log c_\pi(U, A_i)\right] \end{aligned} \tag{124}$$

where, denoting by $(e_i, x_i, y_i) \in \mathbb{R}^{2t+1}$ the i^{th} row of (e, X, Y), we have set

$$\begin{aligned} A_i = A_i(\mathfrak{R}) &= -2U\beta_i^\star + (R, V, W)^\top \\ &\times \left[\frac{1}{n}\begin{pmatrix} e^\top e & e^\top X & e^\top Y \\ X^\top e & X^\top X & X^\top Y \\ Y^\top e & Y^\top X & Y^\top Y \end{pmatrix}\right]^{-1/2} (e_i, x_i, y_i), \\ B_i = B_i(\mathfrak{R}) &= U(\beta_i^\star)^2 - (R, V, W)^\top \\ &\times \left[\frac{1}{n}\begin{pmatrix} e^\top e & e^\top X & e^\top Y \\ X^\top e & X^\top X & X^\top Y \\ Y^\top e & Y^\top X & Y^\top Y \end{pmatrix}\right]^{-1/2} (e_i, x_i, y_i) \cdot \beta_i^\star . \end{aligned}$$

By Theorem 2.2, Corollary 1, and Propositions 10 and 11, almost surely as $n, m \to \infty$, the empirical distributions of $(A_i)_{i=1}^n$ and $(B_i)_{i=1}^n$ converge (in Wasserstein-$\mathfrak{p}$ for every fixed order $\mathfrak{p} \geq 1$) respectively to

$$\begin{aligned} \mathsf{A}(\mathfrak{R}) &= -2U\mathsf{X}^\star + b_*^{-1/2} R\mathsf{E} + V^\top \Delta^{-1/2}(\mathsf{X}_1, \ldots, \mathsf{X}_t) \\ &\quad + \kappa_*^{-1/2} W^\top \Delta^{-1/2}(\mathsf{Y}_1, \ldots, \mathsf{Y}_t), \\ \mathsf{B}(\mathfrak{R}) &= U\mathsf{X}^{\star 2} - b_*^{-1/2} R\mathsf{E}\mathsf{X}^\star - V^\top \Delta^{-1/2}(\mathsf{X}_1, \ldots, \mathsf{X}_t)\mathsf{X}^\star \\ &\quad - \kappa_*^{-1/2} W^\top \Delta^{-1/2}(\mathsf{Y}_1, \ldots, \mathsf{Y}_t)\mathsf{X}^\star. \end{aligned}$$

In the remainder of the proof, we restrict to the event of probability 1 where this empirical convergence holds. These limiting random variables satisfy $\mathbb{E}\mathsf{X}^{\star 2} = \rho_*$, $\mathbb{E}\mathsf{E}\mathsf{X}^\star = 0$, $\mathbb{E}\mathsf{Y}_r\mathsf{X}^\star = 0$, and $\mathbb{E}\mathsf{X}_r\mathsf{X}^\star = \mathbb{E}[F(\mathsf{P}, \mathsf{X}^\star)\mathsf{X}^\star] = \pi_*$, where $\mathsf{P} \sim N(0, \gamma_*^{-1})$ is independent of $\mathsf{X}^\star$, and π_* is defined in (31). Define the limit cumulant generating function

$$\begin{aligned} \lambda(\mathfrak{R}) = &\mathbb{E} \log c_\pi(U, \mathsf{A}(\mathfrak{R})) + \rho_* U \\ &- \pi_* V^\top \Delta^{-1/2} 1_{t\times 1} \in (-\infty, \infty]. \end{aligned}$$

Then Lemma 9 and the above empirical Wasserstein convergence ensure

$$\begin{aligned} \lim_{n,m\to\infty} \lambda_n(\mathfrak{R}) &= \lambda(\mathfrak{R}) < \infty \text{ for all } U < \bar{a}, \\ \lambda_n(\mathfrak{R}) &= \lambda(\mathfrak{R}) = \infty \text{ for all } U > \bar{a}. \end{aligned} \tag{125}$$

Denote the Fenchel-Legendre transform of λ by

$$\lambda^*(\mathfrak{P}) = \sup_{\mathfrak{R} \in \mathbb{R}^{2t+2}} \mathfrak{P}^\top \mathfrak{R} - \lambda(\mathfrak{R}) \in [0, \infty]. \tag{126}$$

Observe that the concentration bound (5) implies that $\int e^{ax^2} d\pi(x) < \infty$ for some sufficiently small $a > 0$. Then $\bar{a} > 0$ strictly, so λ is finite in an open neighborhood of $\mathfrak{R} = 0$, and hence λ^* is a good convex rate function (i.e. lower semi-continuous and having compact level sets) [47, Lemma 2.3.9(a)]. Let us show the large-deviations upper bound

$$\begin{aligned} &\limsup_{n,m\to\infty} \frac{1}{n} \log \left\langle \mathbb{I}\{\mathfrak{P}(\tilde{\sigma}) \in F\} \right\rangle_\pi \\ &\qquad \leq -\inf_{\mathfrak{P} \in F} \lambda^*(\mathfrak{P}) \text{ for all closed } F \subseteq \bar{\mathcal{V}}. \end{aligned} \tag{127}$$

For this, set $\bar{\lambda}(\mathfrak{R}) = \limsup_{n,m\to\infty} \lambda_n(\mathfrak{R})$. Here $\bar{\lambda}(\mathfrak{R}) = \lambda(\mathfrak{R})$ whenever $U \neq \bar{a}$, by (125). The upper bound in the Gärtner-Ellis Theorem shows that (127) holds with λ^* replaced by the Fenchel-Legendre transform $\bar{\lambda}^*$ of $\bar{\lambda}$ (see e.g. [47, Exercise 2.3.25]). Note that both λ and $\bar{\lambda}$ are convex, so the restrictions of λ and $\bar{\lambda}$ to any line segment are upper-semicontinuous by the Gale-Klee-Rockafellar Theorem [48, Theorem 10.2]. Then the supremum in (126) and the analogous supremum defining $\bar{\lambda}^*$ may be restricted to $\{\mathfrak{R} : U \neq \bar{a}\}$, implying that $\lambda^* = \bar{\lambda}^*$. This proves (127).

For the large deviations lower bound, set

$$\lambda^M(\mathfrak{R}) = \mathbb{E} \log c_\pi^M(U, \mathsf{A}(\mathfrak{R})) + \rho_* U - \pi_* V^\top \Delta^{-1/2} 1_{t\times 1},$$

and let $(\lambda^M)^*$ be its Fenchel-Legendre transform defined analogously to (126). We aim to show

$$\begin{aligned} &\liminf_{n,m\to\infty} \frac{1}{n} \log \left\langle \mathbb{I}\{\mathfrak{P}(\tilde{\sigma}) \in G\} \right\rangle_\pi \\ &\quad \geq \sup_{M>0} \left(-\inf_{\mathfrak{P} \in G} (\lambda^M)^*(\mathfrak{P})\right) \text{ for all open } G \subseteq \mathcal{V}. \end{aligned} \tag{128}$$

For this, consider any $M > 0$ where $(-M, M)$ intersects the support of π, and denote by π_M the conditional distribution of π over $(-M, M)$. Let $\langle\cdot\rangle_{\pi_M}$ be the expectation over $(\sigma_i)_{i=1}^n \overset{iid}{\sim} \pi_M$. Then analogously to (124), we have

$$\lambda_n^M(\mathfrak{R}) := \frac{1}{n}\log\Big\langle \exp\big(n\cdot\mathfrak{P}(\tilde\sigma)^\top\mathfrak{R}\big)\Big\rangle_{\pi_M}$$
$$= \frac{1}{n}\sum_{i=1}^n\Big[B_i + \log c_\pi^M(U, A_i)\Big] - \log\pi((-M,M)).$$

By the above empirical Wasserstein convergence, $\lambda_n^M(\mathfrak{R})$ converges pointwise over all $\mathfrak{R}\in\mathbb{R}^{2t+2}$ to $\lambda^M(\mathfrak{R}) - \log\pi((-M,M))$, which is now finite. Then the Gärtner-Ellis lower bound [47, Theorem 2.3.6] may be applied for the law of $\mathfrak{P}(\tilde\sigma)$ under π_M, giving for all open $G\subseteq\mathcal{V}$

$$\frac{1}{n}\log\Big\langle\mathbb{I}\{\mathfrak{P}(\tilde\sigma)\in G\}\Big\rangle_\pi$$
$$\geq \log\pi((-M,M)) + \frac{1}{n}\log\Big\langle\mathbb{I}\{\mathfrak{P}(\tilde\sigma)\in G\Big\rangle_{\pi_M}$$
$$\geq \log\pi((-M,M)) - \inf_{\mathfrak{P}\in G}\big(\lambda^M - \log\pi((-M,M))\big)^*(\mathfrak{P})$$
$$= -\inf_{\mathfrak{P}\in G}(\lambda^M)^*(\mathfrak{P}).$$

This lower bound is increasing in M, and taking the supremum over $M>0$ yields (128).

The function $\mathfrak{P}\mapsto f(\mathfrak{P})$ is continuous and thus bounded over the compact set $\{\mathfrak{P}\in\bar{\mathcal{V}} : u\in[0,K]\}$. Then for any $M>0$, applying (128) and Varadhan's lemma in the form of Lemma 14(a),

$$\liminf_{n,m\to\infty}\frac{1}{n}\log\Big\langle\mathbb{I}\Big\{\mathfrak{P}(\tilde\sigma)\in\mathcal{V},\, u(\tilde\sigma)\in\mathcal{U}\Big\}\exp\Big(\frac{n}{2}f(\mathfrak{P}(\tilde\sigma))\Big)\Big\rangle_\pi$$
$$\geq \sup_{\mathfrak{P}\in\mathcal{V}:\,u\in\mathcal{U}}\frac{1}{2}f(\mathfrak{P}) - (\lambda^M)^*(\mathfrak{P}).$$

Recalling the definitions of $\Phi_{1,t}^M$ and $\Psi_{1,t}^M$ in (34–35), note that $\Psi_{1,t}^M(\mathfrak{P}) = -\frac{1}{2}+\frac{1}{2}f(\mathfrak{P}) - (\lambda^M)^*(\mathfrak{P})$. Then taking the supremum of the above over $M>0$ and applying this to (123) yields the desired lower bound for $\mathbb{E}[\mathcal{Z}(\mathcal{U})\mid\mathcal{G}_t]$. Similarly, recalling that λ^* is a good convex rate function and applying (127) and Varadhan's lemma in the form of Lemma 14(b),

$$\limsup_{n,m\to\infty}\frac{1}{n}\log\Big\langle\mathbb{I}\Big\{\mathfrak{P}(\tilde\sigma)\in\bar{\mathcal{V}},\, u(\tilde\sigma)\in\bar{\mathcal{U}}\Big\}\exp\Big(\frac{n}{2}f(\mathfrak{P}(\tilde\sigma))\Big)\Big\rangle_\pi$$
$$\leq \sup_{\mathfrak{P}\in\bar{\mathcal{V}}:\,u\in\bar{\mathcal{U}}}\frac{1}{2}f(\mathfrak{P}) - \lambda^*(\mathfrak{P}).$$

Note that the condition $\mathfrak{P}(\tilde\sigma)\in\bar{\mathcal{V}}$ on the left side holds always, by definition of $\mathfrak{P}$. Since λ^* is lower-semicontinuous, by the Gale-Klee-Rockafellar Theorem, its restriction to any line segment is continuous. Since f is also continuous, the supremum on the right side may be restricted to the interior $\{\mathfrak{P}\in\mathcal{V} : u\in\mathcal{U}\}$, and this yields the desired upper bound for $\mathbb{E}[\mathcal{Z}(\overline{\mathcal{U}})\mid\mathcal{G}_t]$. ■

We now prove Lemma 2. Let $e_t=(0,\ldots,0,1)$ be the t^{th} standard basis vector in $\mathbb{R}^t$. Denote $\mathfrak{P}_* = (u_*, r_*, v_*, w_*)$ and $\mathfrak{Q}_* = (\zeta_*, U_*, R_*, V_*, W_*, \chi_*^A, \chi_*^B, \chi_*^C)$, where

$$u_* = \frac{2}{\eta_*},\quad r_* = \frac{\gamma_*}{\eta_*}b_*^{1/2},\quad v_* = \frac{\eta_*-\gamma_*}{\eta_*}\Delta_t^{1/2}e_t, \tag{129}$$
$$w_* = \frac{\gamma_*}{\eta_*}\kappa_*^{1/2}\Delta_t^{1/2}e_t,\quad \zeta_* = \eta_*-\gamma_*,$$
$$\chi_*^A = \chi_*^B = \chi_*^C = 0,\quad U_* = -\frac{1}{2}\gamma_*, \tag{130}$$
$$R_* = \gamma_* b_*^{1/2},\quad V_* = 0,\quad W_* = \gamma_*\kappa_*^{1/2}\Delta_t^{1/2}e_t. \tag{131}$$

We will show that $(\mathfrak{P}_*, \mathfrak{Q}_*)$ is an approximate stationary point of $\Phi_{1,t}$, which is an approximate global optimizer of $\sup_{\mathfrak{P}}\inf_{\mathfrak{Q}}\Phi_{1,t}$ for sufficiently small $\mathfrak{e}>0$.

Denote by $\partial_u\Phi_{1,t}\in\mathbb{R}$, $\partial_v\Phi_{1,t}\in\mathbb{R}^t$ etc. the partial derivative or gradient of $\Phi_{1,t}$ in the variables u, v, etc.

Lemma 10: In the setting of Lemma 2, for all $t\geq 1$ and each $\iota\in\{u, r, v, w, \zeta, \chi^A, \chi^B, \chi^C, U, R, W\}$,

$$\Phi_{1,t}(\mathfrak{P}_*, \mathfrak{Q}_*) = \Psi_{\mathrm{RS}},\quad \partial_\iota\Phi_{1,t}(\mathfrak{P}_*, \mathfrak{Q}_*) = 0,$$
$$\lim_{t\to\infty}\|\partial_V\Phi_{1,t}(\mathfrak{P}_*, \mathfrak{Q}_*)\| = 0. \tag{132}$$

Proof: For the first term of $\Phi_{1,t}$, denote

$$\mathsf{A}_* = -2U_*\mathsf{X}^\star + b_*^{-1/2}R_*\mathsf{E} + V_*^\top\Delta_t^{-1/2}(\mathsf{X}_1,\ldots,\mathsf{X}_t) \tag{133}$$
$$+\kappa_*^{-1/2}W_*^\top\Delta_t^{-1/2}(\mathsf{Y}_1,\ldots,\mathsf{Y}_t)$$
$$= \gamma_*(\mathsf{X}^\star + \mathsf{E} + \mathsf{Y}_t). \tag{134}$$

As $\mathsf{E}\sim N(0,b_*)$ and $\mathsf{Y}_t\sim N(0,\sigma_*^2)$ are independent of each other and of $\mathsf{X}^\star$, and $b_*+\sigma_*^2 = \gamma_*^{-1}$, we have

$$\mathbb{E}\log c_\pi(U_*, \mathsf{A}_*) = \mathbb{E}\log c_\pi\left(-\frac{1}{2}\gamma_*, \gamma_*\mathsf{X}^\star + \sqrt{\gamma_*}\mathsf{Z}\right) \tag{135}$$

for $\mathsf{Z}\sim N(0,1)$ independent of $\mathsf{X}^\star$. For the next terms of $\Phi_{1,t}$, we have

$$-(u_*-\rho_*)\cdot U_* - r_*\cdot R_*$$
$$-(v_* + \pi_*\cdot\Delta_t^{-1/2}1_{t\times 1})^\top V_* - w_*^\top W_*$$
$$= \frac{\gamma_*}{2}\left(\frac{2}{\eta_*}-\rho_*\right) - \frac{\gamma_*^2}{\eta_*}b_* - 0 - \frac{\gamma_*^2}{\eta_*}\kappa_*\delta_* = -\frac{1}{2}\gamma_*\rho_* \tag{136}$$

where we used $e_t^\top\Delta_t e_t = \delta_{tt} = \delta_*$ from Theorem 2.2 and the identity $b_*+\kappa_*\delta_* = b_*+\sigma_*^2 = \gamma_*^{-1}$. For the next terms of $\Phi_{1,t}$, applying $e_t^\top\Delta_t e_t = \delta_* = (\eta_*-\gamma_*)^{-1}$ and the definitions of a_*, c_*, e_* in (31), direct calculation gives, after some simplification,

$$\frac{e_*}{b_*}r_*^2 - 2b_*^{-1/2}r_*$$
$$+\operatorname{Tr}\left[\begin{pmatrix} d_* & a_* \\ a_* & c_*\end{pmatrix}(v_*, \kappa_*^{-1/2}w_*)^\top(v_*, \kappa_*^{-1/2}w_*)\right]$$
$$= e_*\left(\frac{\gamma_*}{\eta_*}\right)^2 - 2\frac{\gamma_*}{\eta_*} + d_*\delta_*\left(\frac{\eta_*-\gamma_*}{\eta_*}\right)^2$$
$$+ c_*\delta_*\left(\frac{\gamma_*}{\eta_*}\right)^2 + 2a_*\frac{\gamma_*}{\eta_*^2} = -\frac{\gamma_*}{\eta_*}. \tag{137}$$

Finally, applying again $e_t^\top\Delta_t e_t = \delta_* = (\eta_*-\gamma_*)^{-1}$ and $b_*+\kappa_*\delta_* = \gamma_*^{-1}$, we have

$$\mathcal{A}(\mathfrak{P}_*) = u_* - r_*^2 - \|v_*\|^2 - \|w_*\|^2 = \eta_*^{-1}. \tag{138}$$

We have also $\zeta_* = \eta_* + R(\eta_*^{-1}) = G^{-1}(\eta_*^{-1})$ from (11). Then by [33, Proposition 2.9],

$$\mathcal{H}(\zeta_*, \mathcal{A}(\mathfrak{P}_*)) = \int_0^{\eta_*^{-1}} R(z)dz. \quad (139)$$

The last two terms of $\Phi_{1,t}(\mathfrak{P}_*, \mathfrak{Q}_*)$ are 0 because $\gamma_*\eta_*^{-1} - b_*^{-1/2}r_* = 0$ and $\mathcal{B}(v_*, w_*) = \|v_* - \alpha_*^A w_*\|^2 = 0$. Then combining (135), (136), (137), (139) shows $\Phi_{1,t}(\mathfrak{P}_*, \mathfrak{Q}_*) = \Psi_{\mathrm{RS}}$ in (132).

To check the stationarity conditions, first by the form of $\mathcal{H}$ in (33), we have $\partial_\zeta \mathcal{H}(\zeta, \mathcal{A}) = \mathcal{A} - G(\zeta)$ and $\partial_{\mathcal{A}}\mathcal{H}(\zeta, \mathcal{A}) = \zeta - 1/\mathcal{A}$. Recalling $\zeta_* = \eta_* - \gamma_* = G^{-1}(\eta_*^{-1})$ and $\mathcal{A}(\mathfrak{P}_*) = \eta_*^{-1}$, we have

$$\partial_\zeta \mathcal{H}(\zeta_*, \mathcal{A}(\mathfrak{P}_*)) = 0, \quad \partial_\alpha \mathcal{H}(\zeta_*, \mathcal{A}(\mathfrak{P}_*)) = -\gamma_*. \quad (140)$$

In addition, note that since $[\gamma_*\eta_*^{-1} - b_*^{-1/2}r_*]\mathsf{D}^2 - \chi_*^A = 0$,

$$\partial_r, \partial_\zeta, \partial_{\chi^A}\mathbb{E}\Bigg[\frac{\mathsf{E}_b^2}{\zeta + \mathsf{D}^2} \times \left(\left[\frac{\gamma_*}{\eta_*} - b_*^{-1/2}r\right]\mathsf{D}^2 - \chi^A\right)^2\Bigg]\Bigg|_{r=r_*, \chi^A=\chi_*^A, \zeta=\zeta_*} = 0 \quad (141)$$

Using this and $\mathcal{B}(v_*, w_*), \partial_v\mathcal{B}(v_*, w_*), \partial_w\mathcal{B}(v_*, w_*) = 0$, we obtain

$$\begin{aligned}
\partial_u\Phi_{1,t}(\mathfrak{P}_*, \mathfrak{Q}_*) &= -U_* - \gamma_*/2 = 0\\
\partial_r\Phi_{1,t}(\mathfrak{P}_*, \mathfrak{Q}_*) &= -R_* - (e_*b_*^{-1}r_* - b_*^{-1/2}) + \gamma_* r_*\\
&\overset{(a)}{=} -r_*(\eta_* - \gamma_*) + b_*^{1/2}\frac{\gamma_*(\eta_* - \gamma_*)}{\eta_*} = 0\\
\partial_v\Phi_{1,t}(\mathfrak{P}_*, \mathfrak{Q}_*) &= -V_* - (d_*v_* + a_*\kappa_*^{-1/2}w_*) + \gamma_* v_*\\
&= \big(-d_* - (1 - d_*\gamma_*^{-1})\gamma_* + \gamma_*\big)v_* = 0\\
\partial_w\Phi_{1,t}(\mathfrak{P}_*, \mathfrak{Q}_*) &= -W_* - (c_*\kappa_*^{-1}w_* + a_*\kappa_*^{-1/2}v_*) + \gamma_* w_*\\
&= \left(-\eta_* - \kappa_*^{-1}\left(c_* + \frac{\eta_* - \gamma_*}{\gamma_*}a_*\right) + \gamma_*\right)w_*\\
&\overset{(b)}{=} 0\\
\partial_\zeta, \partial_{\chi^A}, \partial_{\chi^B}, &\partial_{\chi^C}\Phi_{1,t}(\mathfrak{P}_*, \mathfrak{Q}_*) = 0
\end{aligned}$$

where we used in (a) $R_* = \eta_* r_*$ and (by the definition of b_*, e_* in (25) and (31))

$$\begin{aligned}
-e_*b_*^{-1}r_* + b_*^{-1/2} &= b_*^{-1/2}\left(1 - \frac{\gamma_*}{\eta_*}(1 + \kappa_*)\right)\\
&= b_*^{1/2}\frac{\gamma_*(\eta_* - \gamma_*)}{\eta_*},
\end{aligned} \quad (142)$$

and in (b) (by the definitions of a_*, c_* in (31)) $c_* + a_*(\eta_* - \gamma_*)/\gamma_* = -(\eta_* - \gamma_*)\kappa_*$.

Now note that

$$\begin{aligned}
\partial_a \log c_\pi(a, b) &= \frac{\int x^2 \exp\left(ax^2 + bx\right) d\pi(x)}{\int \exp\left(ax^2 + bx\right) d\pi(x)},\\
\partial_b \log c_\pi(a, b) &= \frac{\int x \exp\left(ax^2 + bx\right) d\pi(x)}{\int \exp\left(ax^2 + bx\right) d\pi(x)}.
\end{aligned} \quad (143)$$

Recall (134), where $\mathsf{E}+\mathsf{Y}_t \sim N(0, \gamma_*^{-1})$ is independent of $\mathsf{X}^\star$. Then from the expressions for f, f' in (8) and (61), we have that

$$\begin{aligned}
\partial_a \log c_\pi(U_*, \mathsf{A}_*) &= \gamma_*^{-1}f'(\mathsf{X}^\star + \mathsf{E} + \mathsf{Y}_t) + f(\mathsf{X}^\star + \mathsf{E} + \mathsf{Y}_t)^2\\
\partial_b \log c_\pi(U_*, \mathsf{A}_*) &= f(\mathsf{X}^\star + \mathsf{E} + \mathsf{Y}_t).
\end{aligned} \quad (144)$$

Also recall from Theorem 2.2 that $\mathsf{X}_{t+1} = F(\mathsf{Y}_t + \mathsf{E}, \mathsf{X}^\star)$, so by the definition (23),

$$f(\mathsf{X}^\star + \mathsf{E} + \mathsf{Y}_t) = \frac{\eta_* - \gamma_*}{\eta_*}\mathsf{X}_{t+1} + \frac{\gamma_*}{\eta_*}(\mathsf{Y}_t + \mathsf{E}) + \mathsf{X}^\star. \quad (145)$$

Then, denoting

$$\Delta_{t+1} = \begin{pmatrix} \Delta_t & \delta_t \\ \delta_t^\top & \delta_* \end{pmatrix}$$

and applying Corollary 1 and $\mathbb{E}\mathsf{X}_s\mathsf{X}^\star = \pi_*$ from (31), we have

$$\mathbb{E}(\mathsf{X}_1, \ldots, \mathsf{X}_t)f(\mathsf{X}^\star + \mathsf{E} + \mathsf{Y}_t) = \frac{\eta_* - \gamma_*}{\eta_*}\delta_t + \pi_* 1_{t\times 1}.$$

It follows from this, Stein's lemma, and $\mathbb{E}f'(\mathsf{X}^\star + \mathsf{E} + \mathsf{Y}_t) = \gamma_*\mathbb{E}[\mathbb{V}[\mathsf{X}^\star \mid \mathsf{X}^\star + \mathsf{E} + \mathsf{Y}_t]] = \gamma_*/\eta_*$ that

$$\begin{aligned}
\partial_R\Phi_{1,t}(\mathfrak{P}_*, \mathfrak{Q}_*) &= b_*^{-1/2}\mathbb{E}\mathsf{E}f(\mathsf{X}^\star + \mathsf{E} + \mathsf{Y}_t) - r_*\\
&= b_*^{1/2}\mathbb{E}f'(\mathsf{X}^\star + \mathsf{E} + \mathsf{Y}_t) - r_*\\
&= b_*^{1/2}(\gamma_*/\eta_*) - r_* = 0\\
\partial_V\Phi_{1,t}(\mathfrak{P}_*, \mathfrak{Q}_*) &= \Delta_t^{-1/2}\mathbb{E}(\mathsf{X}_1, \ldots, \mathsf{X}_t)f(\mathsf{X}^\star + \mathsf{E} + \mathsf{Y}_t)\\
&\quad - (v_* + \pi_* \cdot \Delta_t^{-1/2}1_{t\times 1})\\
&= \frac{\eta_* - \gamma_*}{\eta_*}\Delta_t^{-1/2}\delta_t - v_*\\
\partial_W\Phi_{1,t}(\mathfrak{P}_*, \mathfrak{Q}_*) &= \kappa_*^{-1/2}\Delta_t^{-1/2}\mathbb{E}(\mathsf{Y}_1, \ldots, \mathsf{Y}_t)\\
&\quad \times f(\mathsf{X}^\star + \mathsf{E} + \mathsf{Y}_t) - w_*\\
&= \kappa_*^{1/2}\Delta_t^{1/2}e_t \cdot \mathbb{E}f'(\mathsf{X}^\star + \mathsf{E} + \mathsf{Y}_t) - w_* = 0.
\end{aligned}$$

From (145) and the identities $\mathbb{E}\mathsf{X}_{t+1}\mathsf{X}^\star = \pi_*$ and $\mathbb{E}\mathsf{X}_{t+1}^2 = \delta_* = (\eta_* - \gamma_*)^{-1}$, we have also

$$\mathbb{E}\mathsf{X}^\star f(\mathsf{X}^\star + \mathsf{E} + \mathsf{Y}_t) = \frac{\eta_* - \gamma_*}{\eta_*}\pi_* + \rho_* \quad (146)$$

$$\begin{aligned}
\mathbb{E}f(\mathsf{X}^\star + \mathsf{E} + \mathsf{Y}_t)^2 &= \left(\frac{\eta_* - \gamma_*}{\eta_*}\right)^2\frac{1}{\eta_* - \gamma_*} + \left(\frac{\gamma_*}{\eta_*}\right)^2\frac{1}{\gamma_*} \quad (147)\\
&\quad + \rho_* + \frac{2(\eta_* - \gamma_*)}{\eta_*}\pi_*\\
&= \frac{1}{\eta_*} + \rho_* + \frac{2(\eta_* - \gamma_*)}{\eta_*}\pi_*. \quad (148)
\end{aligned}$$

Then

$$\begin{aligned}
\partial_U\Phi_{1,t}(\mathfrak{P}_*, \mathfrak{Q}_*) = \mathbb{E}\Big[&\gamma_*^{-1}f'(\mathsf{X}^\star + \mathsf{E} + \mathsf{Y}_t)\\
&+ f(\mathsf{X}^\star + \mathsf{E} + \mathsf{Y}_t)^2\Big]\\
&- 2\mathbb{E}\mathsf{X}^\star f(\mathsf{X}^\star + \mathsf{E} + \mathsf{Y}_t)\\
&- (u_* - \rho_*) = 0.
\end{aligned}$$

This shows $\partial_\iota\Phi_{1,t}(\mathfrak{P}_*, \mathfrak{Q}_*) = 0$ in (132) for all $\iota \neq V$, and it remains to verify the bound for $\partial_V\Phi_{1,t}$. For this, note that from the above and $\delta_{tt} = \delta_{t+1,t+1} = \delta_*$,

$$\begin{aligned}\|\partial_V \Phi_{1,t}(\mathfrak{P}_*, \mathfrak{Q}_*)\|^2 &= \left\| \frac{\eta_* - \gamma_*}{\eta_*} \Delta_t^{-1/2} \delta_t - v_* \right\|^2 \\ &= \left(\frac{\eta_* - \gamma_*}{\eta_*} \right)^2 \left\| \Delta_t^{-1/2} \delta_t - \Delta_t^{1/2} e_t \right\|^2 \\ &= \left(\frac{\eta_* - \gamma_*}{\eta_*} \right)^2 (\delta_t^\top \Delta_t^{-1} \delta_t + \delta_* - 2\delta_t^\top e_t) \\ &= \left(\frac{\eta_* - \gamma_*}{\eta_*} \right)^2 \Big(\delta_t^\top \Delta_t^{-1} \delta_t - \delta_{t+1,t+1} \\ &\quad + 2\delta_* - 2\delta_{t,t+1} \Big).\end{aligned}$$

By Proposition 4, $\lim_{t\to\infty} \delta_{t,t+1} = \delta_*$. By the condition $\Delta_{t+1} \succ 0$ and the Schur-complement formula,

$$\begin{aligned}0 < \delta_{t+1,t+1} - \delta_t^\top \Delta_t^{-1} \delta_t &= \inf_{\alpha\in\mathbb{R}^t} \mathbb{E}\left[(\mathsf{X}_{t+1} - \alpha^\top (\mathsf{X}_1, \ldots, \mathsf{X}_t))^2 \right] \\ &\le \mathbb{E}\left[(\mathsf{X}_{t+1} - \mathsf{X}_t)^2 \right] = 2\delta_* - 2\delta_{t,t+1}.\end{aligned}$$

So, we also have that $\lim_{t\to\infty} \delta_{t+1,t+1} - \delta_t^\top \Delta_t^{-1} \delta_t = 0$. Thus $\lim_{t\to\infty} \|\partial_V \Phi_{1,t}(\mathfrak{P}_*, \mathfrak{Q}_*)\| = 0$. ■

Proof of Lemma 2:

We begin with the lower bound in (36). Denote by $o_t(1)$ any scalar quantity that converges to 0 as $t \to \infty$.

Let us specialize $\Psi_{1,t}^M$ to $\tilde{\mathfrak{P}}_* = (u_*, r_*, \tilde{v}_*, w_*)$ where

$$\tilde{v}_* = \frac{\eta_* - \gamma_*}{\eta_*} \Delta_t^{-1/2} \delta_t = v_* + \partial_V \Phi_{1,t}(\mathfrak{P}_*, \mathfrak{Q}_*). \quad (149)$$

Then Lemma 10 shows $\|\tilde{v}_* - v_*\| = o_t(1)$. Observing that $\|w_*\|^2 = (\gamma_*/\eta_*)^2 \kappa_* \delta_*$ is constant in t, and applying (137), we then have

$$\begin{aligned}\Psi_{1,t}^M(\tilde{\mathfrak{P}}_*) = &-\frac{1}{2} + \frac{\gamma_*}{2\eta_*} + o_t(1) + \inf_{U,R,V,W} X^M(U, R, V, W) \\ &+ \frac{1}{2} \inf_{\zeta,\chi^A,\chi^B,\chi^C} Y(\zeta, \chi^A, \chi^B, \chi^C) \quad (150)\end{aligned}$$

where

$$\begin{aligned}X^M(U, R, V, W) = \mathbb{E} \log c_\pi^M \Big(&U, \\ &- 2U\mathsf{X}^\star + b_*^{-1/2} R\mathsf{E} \\ &+ V^\top \Delta_t^{-1/2} (\mathsf{X}_1, \ldots, \mathsf{X}_t) \\ &+ \kappa_*^{-1/2} W^\top \Delta_t^{-1/2} (\mathsf{Y}_1, \ldots, \mathsf{Y}_t) \Big) \\ &- (u_* - \rho_*)U - r_* R \\ &- (\tilde{v}_* + \pi_* \Delta_t^{-1/2} 1_{t\times 1})^\top V - w_*^\top W,\end{aligned}$$

$$\begin{aligned}Y(\zeta, \chi^A, \chi^B, \chi^C) = \mathcal{H}(\zeta, \mathcal{A}(\tilde{\mathfrak{P}}_*)) &+ \mathbb{E}\left[\frac{\mathsf{E}_b^2}{\zeta + \mathsf{D}^2} (\chi^A)^2 \right] \\ &+ \mathbb{E}\Big[\frac{1}{\zeta + \mathsf{D}^2} \Big(\theta(\mathsf{D}^2) \\ &- \lambda(\mathsf{D}^2) \chi^B - \chi^C \Big)^2 \Big] \mathcal{B}(\tilde{v}_*, w_*).\end{aligned}$$

Let $X(U, R, V, W)$ have the same definition as $X^M(U, R, V, W)$ with c_π^M replaced by c_π. We note that by Lemma 10, (149) is defined exactly so that (U_*, R_*, V_*, W_*) is a critical point of $X(\cdot)$. Since $X(\cdot)$ is convex, (U_*, R_*, V_*, W_*) is then a global minimizer of X. We claim that

$$\sup_{M>0} \inf_{U,R,V,W} X^M(U, R, V, W) = X(U_*, R_*, V_*, W_*). \quad (151)$$

This is trivial if π has bounded support, because $X^M(U, R, V, W)$ is increasing in M and equals $X(U, R, V, W)$ whenever $[-M, M]$ contains the support of π. To show (151) when π has unbounded support, let us check the strict convexity of $X(\cdot)$: Fix any unit vector $(U', R', V', W') \in \mathbb{R}^{2t+2}$ and $s > 0$, and denote

$$\begin{aligned}(U(s), R(s), V(s), W(s)) = &(U_*, R_*, V_*, W_*) \\ &+ s \cdot (U', R', V', W').\end{aligned}$$

Set

$$\mathsf{F} = \begin{pmatrix} b_*^{-1/2} \mathsf{E} \\ \Delta_t^{-1/2} (\mathsf{X}_1, \ldots, \mathsf{X}_t) \\ \kappa_*^{-1/2} \Delta_t^{-1/2} (\mathsf{Y}_1, \ldots, \mathsf{Y}_t) \end{pmatrix} \in \mathbb{R}^{2t+1}. \quad (152)$$

Recalling $U_* = -\gamma_*/2$ and A_* from (134), denote

$$\langle f(x) \rangle_* = \frac{\int f(x) \exp\left(-\frac{\gamma_*}{2} x^2 + \mathsf{A}_* x\right) d\pi(x)}{\int \exp\left(-\frac{\gamma_*}{2} x^2 + \mathsf{A}_* x\right) d\pi(x)}$$

with the corresponding variance $\mathbb{V}_*[f(x)] = \langle f(x)^2 \rangle_* - \langle f(x) \rangle_*^2$ (conditional on $\mathsf{X}^\star, \mathsf{E}, \mathsf{X}_1, \ldots, \mathsf{X}_t, \mathsf{Y}_1, \ldots, \mathsf{Y}_t$). Then, applying (143) and the chain rule,

$$\begin{aligned}&\partial_s^2 X(U(s), R(s), V(s), W(s))\big|_{s=0} \\ &= \mathbb{E}\left[\mathbb{V}_*\left[U'(x - \mathsf{X}^\star)^2 + (R', V', W')^\top \mathsf{F} \cdot x \right] \right]. \quad (153)\end{aligned}$$

Since π has unbounded support, there are at least three distinct points in its support, and hence also three distinct points in the support of the posterior measure defining $\langle\cdot\rangle_*$. Then the conditional variance $\mathbb{V}_*\left[U'(x - \mathsf{X}^\star)^2 + (R', V', W')^\top \mathsf{F} \cdot x \right]$ is 0 only if the quadratic function $x \mapsto U'(x - \mathsf{X}^\star)^2 + (R', V', W')^\top \mathsf{F} \cdot x$ takes constant value at these three points, which occurs only when both $U' = 0$ and $(R', V', W')^\top \mathsf{F} = 0$. When $U' = 0$, we have $\|(R', V', W')\| = 1$. By Theorem 2.2, Corollary 1, and Proposition 7, F has zero mean and identity covariance, so $(R', V', W')^\top \mathsf{F}$ has variance 1. Then in particular, $(R', V', W')^\top \mathsf{F} \ne 0$ with positive probability. Then $\mathbb{V}_*\left[U'(x - \mathsf{X}^\star)^2 + (R', V', W')^\top \mathsf{F} \cdot x \right] > 0$ with positive probability, and hence (153) is strictly positive. This shows the strict convexity $\nabla^2 X(U_*, R_*, V_*, W_*) \succ 0$ as desired. Then by continuity, also $\nabla^2 X(U, R, V, W) \succ 0$ and $X(U, R, V, W) < \infty$ in a bounded neighborhood $\mathcal{O}$ of (U_*, R_*, V_*, W_*). By the monotone convergence theorem, $\lim_{M\to\infty} X^M(U, R, V, W) = X(U, R, V, W)$, and this convergence is uniform over $\mathcal{O}$ because X^M and X are convex [48, Theorem 10.8]. Then the infimum of $X^M(U, R, V, W)$ is attained also in $\mathcal{O}$ for all large M, and $\lim_{M\to\infty} \inf_{U,R,V,W} X^M(U, R, V, W) = X(U_*, R_*, V_*, W_*)$. Hence (151) holds, as claimed.

For the second term $Y(\zeta, \chi^A, \chi^B, \chi^C)$, recall $\mathcal{A}(\mathfrak{P}_*) = \eta_*^{-1}$ from (138), and $\partial_\zeta \mathcal{H}(\zeta_*, \eta_*^{-1}) = 0$. Then, since $\|v_*\|^2 - \|\tilde{v}_*\|^2 = o_t(1)$, we have $\mathcal{A}(\tilde{\mathfrak{P}}_*) = \eta_*^{-1} + o_t(1)$ and $\partial_\zeta \mathcal{H}(\zeta_*, \mathcal{A}(\tilde{\mathfrak{P}}_*)) = o_t(1)$. Furthermore $\partial_\zeta^2 \mathcal{H}(\zeta, \mathcal{A}(\tilde{\mathfrak{P}}_*)) = -G'(\zeta) > 0$ in a neighborhood of ζ_*, so $\inf_{\zeta > d_-} \mathcal{H}(\zeta, \mathcal{A}(\tilde{\mathfrak{P}}_*)) = \mathcal{H}(\zeta_*, \eta_*^{-1}) + o_t(1)$

(see e.g. [33, Proposition C.2]). Applying that the infimum of $Y(\zeta,\chi^A,\chi^B,\chi^C)$ occurs at $\chi^A = 0$, and that $\mathcal{B}(\tilde v_*,w_*) = \|v_* - \tilde v_*\|^2 = o_t(1)$, we obtain

$$\inf_{\zeta,\chi^A,\chi^B,\chi^C} Y(\zeta,\chi^A,\chi^B,\chi^C) = \mathcal{H}(\zeta_*,\eta_*^{-1}) + o_t(1). \quad (154)$$

Applying (151) and (154) to (150), we obtain

$$\lim_{t\to\infty}\sup_{M>0}\Psi_{1,t}^M(\tilde{\mathfrak{P}}_*) = -\frac{1}{2} + \frac{\gamma_*}{2\eta_*} + X(U_*,R_*,V_*,W_*) + \frac{1}{2}\mathcal{H}(\zeta_*,\eta_*^{-1}) = \Psi_{\rm RS},$$

which implies the lower bound in (36).

For the upper bounds in (36) and (37), we now fix $\mathfrak{h}\in\mathbb{R}_+$ and specialize the dual variables $\mathfrak{Q} = \mathfrak{Q}(\mathfrak{P})$ as functions of $\mathfrak{P} = (u,r,v,w)$, given by

$$\begin{aligned} &U(u) = U_* + \mathfrak{h}(u-u_*), \quad R(r) = R_* + \mathfrak{h}(r-r_*),\\ &V(v) = V_* + \mathfrak{h}(v-v_*), \quad W(w) = W_* + \mathfrak{h}(w-w_*),\\ &\zeta(\mathfrak{P}) = G^{-1}(\mathcal{A}(\mathfrak{P})),\\ &\chi^A(r) = -\left(\frac{\gamma_*(\eta_*-\gamma_*)}{\eta_*} - b_*^{-1} + r b_*^{-3/2} e_*\right),\\ &\chi^B = \chi_*^B = 0, \quad \chi^C = \chi_*^C = 0. \end{aligned}$$

Here, for any $\mathfrak{P}\in\mathcal{V}$ with $u\in(0,K)$, we have $\mathcal{A}(\mathfrak{P}) < K$ so $\zeta(\mathfrak{P})$ is well-defined for all $\mathfrak{e}$ sufficiently small. Note that under the above choices of χ^A,χ^B,χ^C, we have

$$\mathcal{F}_{22}^e(\zeta,r) = \mathbb{E}\left[\frac{\mathsf{E}_b^2}{\zeta+\mathsf{D}^2}\left(\left[\frac{\gamma_*}{\eta_*} - b_*^{-1/2} r\right]\mathsf{D}^2 - \chi^A(r)\right)^2\right],$$

$$\mathcal{F}_{22}(\zeta) = \mathbb{E}\left[\frac{1}{\zeta+\mathsf{D}^2}\left(\theta\left(\mathsf{D}^2\right) - \lambda\left(\mathsf{D}^2\right)\chi_*^B - \chi_*^C\right)^2\right]$$

where $\mathcal{F}_{22}^e,\mathcal{F}_{22}$ are the functions previously defined in (110). Then

$$\begin{aligned}\Psi_{1,t}(\mathfrak{P}) &= \inf_{\mathfrak{Q}}\Phi_{1,t}(\mathfrak{P},\mathfrak{Q}) \le \Phi_{1,t}(\mathfrak{P},\mathfrak{Q}(\mathfrak{P})) =: \bar\Psi_{1,t}(\mathfrak{P})\\ &= -\frac{1}{2} + \mathrm{I} + \mathrm{II} + \frac{1}{2}\Big(\mathrm{III}+\mathrm{IV}+\mathrm{V}+\mathrm{VI}\Big) \quad (155)\end{aligned}$$

where, recalling F from (152), we define

$$\begin{aligned} \mathrm{I} &= \mathbb{E}\log c_\pi\Big(U(u), -2U(u)\mathsf{X}^\star + (R(r),V(v),W(w))^\top\mathsf{F}\Big)\\ \mathrm{II} &= -(u-\rho_*)\cdot U(u) - r\cdot R(r)\\ &\quad - (v + \pi_*\Delta_t^{-1/2}1_{t\times 1})^\top V(v) - w^\top W(w)\\ \mathrm{III} &= -\frac{e_* r^2}{b_*} + \frac{2r}{\sqrt{b_*}} - \mathrm{Tr}\begin{pmatrix} d_* & a_*\\ a_* & c_*\end{pmatrix}\left(v,\frac{w}{\sqrt{\kappa_*}}\right)^\top\left(v,\frac{w}{\sqrt{\kappa_*}}\right)\\ \mathrm{IV} &= \mathcal{H}\left(\zeta(\mathfrak{P}),\mathcal{A}(\mathfrak{P})\right), \quad \mathrm{V} = \mathcal{F}_{22}^e(\zeta(\mathfrak{P}),r),\\ \mathrm{VI} &= \mathcal{F}_{22}(\zeta(\mathfrak{P}))\cdot\mathcal{B}(v,w) \end{aligned} \quad (156)$$

At $\mathfrak{P} = \mathfrak{P}_*$, using $\mathcal{A}(\mathfrak{P}_*) = \eta_*^{-1}$, $\zeta_* = G^{-1}(\eta_*^{-1})$, and (142) to verify $\chi^A(r_*) = 0$, we observe that these specializations give $\mathfrak{Q}(\mathfrak{P}_*) = \mathfrak{Q}_*$. Then $\bar\Psi_{1,t}(\mathfrak{P}_*) = \Phi_{1,t}(\mathfrak{P}_*,\mathfrak{Q}_*) = \Psi_{\rm RS}$ by Lemma 10. Furthermore, noting that the only coordinates of $\mathfrak{Q}(\mathfrak{P})$ depending on v are $V(v)$ and $\zeta(\mathfrak{P})$, the derivative of $\bar\Psi_{1,t}$ in v is

$$\partial_v\bar\Psi_{1,t}(\mathfrak{P}_*) = \partial_v\Phi_{1,t}(\mathfrak{P}_*,\mathfrak{Q}_*) + \partial_\zeta\Phi_{1,t}(\mathfrak{P}_*,\mathfrak{Q}_*)\cdot\partial_v\zeta(\mathfrak{P}_*) + \partial_V\Phi_{1,t}(\mathfrak{P}_*,\mathfrak{Q}_*)\cdot\partial_v V(v_*).$$

The first term has norm $o_t(1)$, and the remaining two terms are 0, by Lemma 10. Similarly $\partial_u\bar\Psi_{1,t}(\mathfrak{P}_*) = 0$, $\partial_r\bar\Psi_{1,t}(\mathfrak{P}_*) = 0$, and $\partial_w\bar\Psi_{1,t}(\mathfrak{P}_*) = 0$, so $\|\nabla\bar\Psi_{1,t}(\mathfrak{P}_*)\| = o_t(1)$.

We now show, using the small-$\mathfrak{e}$ approximations of Proposition 8, that the upper bound $\bar\Psi_{1,t}(\mathfrak{P})$ in (155) is concave in $\mathfrak{P}$ over the domain $\{\mathfrak{P}\in\mathcal{V} : u\in(0,K)\}$. Let us write $O(1)$, $O(\mathfrak{e})$ etc. for scalar quantities bounded in magnitude by C, $C\mathfrak{e}$, etc. where the constant $C>0$ depends only on $\mathfrak{C},K$ (and not on $d_*,\mathfrak{e},\mathfrak{h}$ or the dimension t). Fix any $\mathfrak{P} = (u,r,v,w)\in\mathcal{V}$ with $u\in(0,K)$, fix any unit vector $(u',r',v',w')\in\mathbb{R}^{2t+2}$, and define for $s>0$

$$\mathfrak{P}(s) = (u(s),r(s),v(s),w(s)) = (u,r,v,w) + s\cdot(u',r',v',w'). \quad (157)$$

We compute the second derivative of $\bar\Psi_{1,t}(\mathfrak{P}(s))$ at $s=0$. For the first term I, denote

$$\langle f(x)\rangle_{\mathfrak{P}} = \frac{A}{B}$$

where

$$A := \int f(x)\exp\Big(U(u)\cdot x^2 - 2U(u)\mathsf{X}^\star\cdot x + (R(r),V(v),W(w))^\top\mathsf{F}\cdot x\Big)d\pi(x)$$

$$B := \int \exp\Big(U(u)\cdot x^2 - 2U(u)\mathsf{X}^\star\cdot x + (R(r),V(v),W(w))^\top\mathsf{F}\cdot x\Big)d\pi(x)$$

and let $\mathbb{V}_{\mathfrak{P}}[f(x)] = \langle f(x)^2\rangle_{\mathfrak{P}} - \langle f(x)\rangle_{\mathfrak{P}}^2$ be the corresponding variance. Then

$$\begin{aligned}\partial_s^2\mathrm{I}\big|_{s=0} &= \mathfrak{h}^2\,\mathbb{E}\left[\mathbb{V}_{\mathfrak{P}}\left[u'x^2 - 2u'\mathsf{X}^\star\cdot x + (r',v',w')^\top\mathsf{F}\cdot x\right]\right]\\ &\le 2\mathfrak{h}^2\,\mathbb{E}\Big[(u')^2\cdot\mathbb{V}_{\mathfrak{P}}[x^2] \quad (158)\\ &\quad + \left(-2u'\mathsf{X}^\star + (r',v',w')^\top\mathsf{F}\right)^2\mathbb{V}_{\mathfrak{P}}[x]\Big]. \quad (159)\end{aligned}$$

Let us apply Assumption 1.3 and Proposition 16 in dimension $k=1$, with $\Gamma = \gamma_{\max} = \gamma_{\min} = U(u)$ and $z = -2U(u)\mathsf{X}^\star + (R(r),V(v),W(w))^\top\mathsf{F}$. We observe that, since $u,u_*\in(0,K)$, we have

$$\begin{aligned}U(u) = U_* + \mathfrak{h}(u-u_*) &= -(\gamma_*/2) + \mathfrak{h}(u-u_*)\\ &= -(d_*/2)(1+O(\mathfrak{e})) + O(\mathfrak{h}),\end{aligned} \quad (160)$$

the last equality applying Proposition 8. In particular, since $d_*>0$, for all $\mathfrak{h}\in(0,\mathfrak{h}_0)$ where $\mathfrak{h}_0$ is a small constant depending only on $(K,\mathfrak{C})$, we have $\Gamma < (4\mathfrak{C})^{-1}$. Then the condition (6) from Assumption 1.3 implies $\mathbb{V}_{\mathfrak{P}}[x] = O(1)$. Since also $r^2 + \|v\|^2 + \|w\|^2 < u < K$ and $r_*^2 + \|v_*\|^2 + \|w_*\|^2 < u_* < K$ and $b_* + \kappa_*\delta_* = \gamma_*^{-1}$ and $\gamma_* = d_*(1+O(\mathfrak{e}))$, we have

$$\begin{aligned}&\|(R(r),V(v),W(w))\| \le \|(R_*,V_*,W_*)\| + O(\mathfrak{h})\\ &= (\gamma_*^2 b_* + \gamma_*^2\kappa_*\delta_*)^{1/2} + O(\mathfrak{h}) = d_*^{1/2}(1+O(\mathfrak{e})) + O(\mathfrak{h}).\end{aligned}$$

Then, for all $\mathfrak{h}\in(0,\mathfrak{h}_0)$, we have $\|z\|^2 \le (\mathsf{X}^{\star 2} + (q^\top\mathsf{F})^2)\cdot O(1+d_*)$ for some unit vector $q\in\mathbb{R}^{2t+1}$. From (160), we have $\mathfrak{C}^{-1} - \gamma_{\max} \geq c(1+d_*)$ and $-\gamma_{\min} < C(1+d_*)$ for constants $C, c > 0$ depending only on $(K, \mathfrak{C})$. So Proposition 16 shows $\mathbb{V}_{\mathfrak{P}}[x^2] \leq C(1+\mathsf{X}^{\star 2}+(q^\top \mathsf{F})^2)$. Applying these to (159), and applying also $\mathbb{E}[\mathsf{X}^{\star 2}] \leq \mathfrak{C}$ and $\mathbb{E}[(q^\top \mathsf{F})^2] = 1$ for any unit vector q, we obtain

$$\partial_s^2 \mathrm{I}\big|_{s=0} = O(\mathfrak{h}^2). \tag{161}$$

For II, we have

$$\partial_s^2 \mathrm{II}\ |_{s=0} = -2\mathfrak{h}({u'}^2 + {r'}^2 + \|v'\|^2 + \|w'\|^2) = -2\mathfrak{h}. \tag{162}$$

For III, applying Proposition 8,

$$\partial_s^2 \mathrm{III}\big|_{s=0} = -2\mathrm{Tr}\begin{pmatrix} d_* & a_*\kappa_*^{-1/2} \\ a_*\kappa_*^{-1/2} & c_*\kappa_*^{-1}\end{pmatrix}(v', w')^\top (v', w') - \frac{2e_*}{b_*}{r'}^2 = -2d_*\|(r', v', w')\|^2 + O(\mathfrak{e}). \tag{163}$$

For IV, we have $\mathrm{IV} = \int_0^{\mathcal{A}(\mathfrak{P}(s))} R(z)dz$ by [33, Proposition 2.9(a)]. It is easily checked that $|\mathcal{A}(\mathfrak{P}(s))| = O(1)$, $|\partial_s \mathcal{A}(\mathfrak{P}(s))| = O(1)$, and $\partial_s^2 \mathcal{A}(\mathfrak{P}(s)) = -2\|(r', v', w')\|^2$ at $s = 0$. Then by Proposition 8,

$$\begin{aligned}\partial_s^2 \mathrm{IV}\big|_{s=0} &= R'(\mathcal{A}(\mathfrak{P}(s))) \cdot \partial_s \mathcal{A}(\mathfrak{P}(s))^2 \\ &\quad + R(\mathcal{A}(\mathfrak{P}(s))) \cdot \partial_s^2 \mathcal{A}(\mathfrak{P}(s))^2\Big|_{s=0} \\ &= 2d_*\|(r', v', w')\|^2 + O(\mathfrak{e}^2).\end{aligned} \tag{164}$$

For V, we may apply the series expansion for $R(z)$ from (219) with $\kappa_1 = -\mathbb{E}[\mathsf{D}^2] = -d_*$, to write for any $z \in (0, K)$, $x \in [d_* - \mathfrak{e}, d_* + \mathfrak{e}]$, and sufficiently small $\mathfrak{e}$,

$$(G^{-1}(z) + x)^{-1} = \big(R(z) + z^{-1} + x\big)^{-1} \tag{165}$$

$$= z\left(1 + (x - d_*)z + \sum_{k\geq 2}\kappa_k z^k\right)^{-1}$$

$$= z \cdot \sum_{j\geq 0}\left(-(x-d_*)\,z - \sum_{k\geq 2}\kappa_k z^k\right)^j \tag{166}$$

$$=: z + \sum_{k\geq 1} c_k(x) z^{k+1}. \tag{167}$$

Here, $|c_k(x)| \leq (O(\mathfrak{e}))^k$, and these series are absolutely convergent for sufficiently small $\mathfrak{e}$. Then the derivatives in z may be computed term-by-term. Recalling $\zeta(\mathfrak{P}(s)) = G^{-1}(\mathcal{A}(\mathfrak{P}(s)))$ where $\mathcal{A}(\mathfrak{P}(0)) \in (0, K)$, we obtain by the chain rule

$$\sup_{x\in\mathrm{supp}(\mathsf{D}^2)} \left|\partial_s^k(\zeta(\mathfrak{P}(s)) + x)^{-1}\right|\Big|_{s=0} = O(1) \text{ for } k = 0, 1, 2. \tag{168}$$

Recalling $\upsilon^A(x)$ from (110) and applying Proposition 8, we have for any $x \in [d_* - \mathfrak{e}, d_* + \mathfrak{e}]$,

$$\upsilon^A(x) = b_*^{-1/2}\left(-x + \frac{e_*}{b_*}\right) = O(b_*^{-1/2}\mathfrak{e}).$$

By Proposition 8, we have $\gamma_* = d_* + O(\kappa_2/\eta_*)$ and also $b_*^{-1} = d_*(1 + \eta_*^{-1} \cdot O(\kappa_2/d_*)) = d_* + O(\kappa_2/\eta_*)$. Then recalling $\upsilon^B(x)$ from (110), we have

$$\begin{aligned}\upsilon^B(x) &= \left(\frac{\gamma_*}{\eta_*}\right)x + \left(\frac{\gamma_*\,(\eta_* - \gamma_*)}{\eta_*} - \frac{1}{b_*}\right) \\ &= \frac{\gamma_*}{\eta_*}(x - \gamma_*) + O\left(\frac{\kappa_2}{\eta_*}\right) \\ &= O\left(\frac{\gamma_*\mathfrak{e}}{\eta_*} + \frac{\kappa_2}{\eta_*}\right).\end{aligned}$$

Here, $\kappa_2 = O(d_*\mathfrak{e}) = O(\gamma_*\mathfrak{e})$, the second equality holding because $\gamma_* = d_*(1 + O(\mathfrak{e}))$. Then, applying also $\eta_*^{-1} = O(1)$ and $0 < \gamma_* < \eta_*$ by Proposition 1, this gives

$$\begin{aligned}\upsilon^B(x) &= O\left(\frac{\gamma_*\mathfrak{e}}{\eta_*}\right) \\ &= O\left(\sqrt{\frac{\gamma_*}{\eta_*}} \cdot \sqrt{\frac{1}{\eta_*}} \cdot \sqrt{\gamma_*} \cdot e\right) \\ &= O(\sqrt{d_*} \cdot \mathfrak{e}) \\ &= O(b_*^{-1/2} \cdot \mathfrak{e}).\end{aligned}$$

Now applying these bounds for υ^A, υ^B to $\mathfrak{f}^e(x, r)$ from (110) and differentiating by the chain rule,

$$\sup_{x\in\mathrm{supp}(\mathsf{D}^2)} \left|\partial_s^k \mathfrak{f}^e(x, r(s))\right|\Big|_{s=0} = O(b_*^{-1/2} \cdot \mathfrak{e}) \text{ for } k = 0, 1, 2. \tag{169}$$

Combining (168) and (169) and differentiating $\mathcal{F}_{22}^e$ from (110) by the chain rule,

$$\begin{aligned}\partial_s^2 \mathrm{V}\big|_{s=0} &= \partial_s^2 \mathcal{F}_{22}^e(\zeta(\mathfrak{P}(s)), r(s))\Big|_{s=0} \\ &\leq O(b_*^{-1} \cdot \mathfrak{e}^2) \cdot \mathbb{E}[\mathsf{E}_b^2] \\ &= O(b_*^{-1} \cdot \mathfrak{e}^2) \cdot b_* \\ &= O(\mathfrak{e}^2).\end{aligned} \tag{170}$$

For VI, for any $x \in [d_* - \mathfrak{e}, d_* + \mathfrak{e}]$, we may write $\theta(x)$ from (33) as

$$\begin{aligned}\theta(x) &= x - d_* - \alpha_*^B\left(\frac{1}{1 - \eta_*^{-1}(\gamma_* - x)} - 1\right) \\ &= x - d_* - \alpha_*^B \sum_{k\geq 1}\left(\frac{\gamma_* - x}{\eta_*}\right)^k \\ &= (x - \gamma_*)\left(1 + \eta_*^{-1}\alpha_*^B\right) + \gamma_* \\ &\quad - d_* - \alpha_*^B\sum_{k\geq 2}\left(\frac{\gamma_* - x}{\eta_*}\right)^k.\end{aligned}$$

Then, using $|x - \gamma_*|/\eta_* = O(\mathfrak{e})$, we have

$$\mathbb{E}\left[\theta(\mathsf{D}^2)^2\right] \leq 3\Bigg((1 + \eta_*^{-1}\alpha_*^B)^2 \cdot \mathbb{E}\left[(\mathsf{D}^2 - \gamma_*)^2\right] + (\gamma_* - d_*)^2 + \left(\alpha_*^B\right)^2 \mathbb{E}\left[\left(\frac{\gamma_* - \mathsf{D}^2}{\eta_*}\right)^4\right](1 + O(\mathfrak{e}))\Bigg)$$

By Proposition 8 and Lemma 13(c), we have $(\gamma_* - d_*)^2 = O(\kappa_2^2\eta_*^{-2}) = O(\kappa_2\mathfrak{e}^2/\eta_*^2)$, $\mathbb{E}[(\mathsf{D}^2 - \gamma_*)^2] = \kappa_2 + (d_* - \gamma_*)^2 =$

$O(\kappa_2)$, $\mathbb{E}[(\mathsf{D}^2-\gamma_*)^4] = O(\mu_4 + (d_* - \gamma_*)^4) = O(\kappa_2 \mathfrak{e}^2)$, $1+\eta_*^{-1}\alpha_*^B = O(\mathfrak{e}/\eta_*)$, and $\eta_*^{-2}(\alpha_*^B)^2 = O(1)$. This gives

$$\mathbb{E}[\theta(\mathsf{D}^2)^2] = O(\kappa_2 \mathfrak{e}^2/\eta_*^2) \tag{171}$$

Then, applying again (168) and differentiating $\mathcal{F}_{22}$ from (110), we have at $s=0$ that $\partial_s^k \mathcal{F}_{22}(\zeta(\mathfrak{P}(s))) = O(\kappa_2 \mathfrak{e}^2/\eta_*^2)$ for $k=0,1,2$. Using $\alpha_*^A = \frac{\eta_*}{\sqrt{\kappa_2}}(1+O(\eta_*^{-1}\mathfrak{e}))$ and $\|v\|, \|w\|, \|v'\|, \|w'\| = O(1)$, we have also $\partial_s^k \mathcal{B}(v(s), w(s)) = O(\max(1, \eta_*^2/\kappa_2))$ for $k=0,1,2$. Combining these bounds, we conclude that

$$\partial_s^2 \mathrm{VI}\Big|_{s=0} = O(\mathfrak{e}^2). \tag{172}$$

Now, combining (161), (162), (163), (164), (170), (172), and setting $\mathfrak{h} = \mathfrak{e}^{1/2}$, we conclude that

$$\partial_s^2 \bar{\Psi}_{1,t}(\mathfrak{P}(s))|_{s=0} = -2\mathfrak{e}^{1/2} + O(\mathfrak{e}) < -\mathfrak{e}^{1/2}$$

for all $\mathfrak{e} < \mathfrak{e}_0(\mathfrak{C}, K)$. This holds for $\mathfrak{P}(s)$ as defined in (157) for any $(u,r,v,w) \in \mathcal{V}$ with $u \in (0,K)$ and for any unit vector (u',r',v',w'), implying the concavity

$$\nabla^2 \bar{\Psi}_{1,t}(\mathfrak{P}) \prec -\mathfrak{e}^{1/2} I \text{ over } \{\mathfrak{P} \in \mathcal{V} : u \in (0,K)\}. \tag{173}$$

Finally, since $u_* = 2\eta_*^{-1} \in \mathcal{U} \subseteq (0,K)$ by assumption, we have that $\mathfrak{P}_* = (u_*, r_*, v_*, w_*)$ is an interior point of the open domain $\{\mathfrak{P} \in \mathcal{V} : u \in \mathcal{U}\}$. Recalling $\bar{\Psi}_{1,t}(\mathfrak{P}_*) = \Psi_{\mathrm{RS}}$ and $\|\nabla \bar{\Psi}_{1,t}(\mathfrak{P}_*)\| = o_t(1)$, we then have (see e.g. [33, Proposition C.2])

$$\begin{aligned}\sup_{\mathfrak{P}\in\mathcal{V}:u\in\mathcal{U}} \Psi_{1,t}(\mathfrak{P}) &\le \sup_{\mathfrak{P}\in\mathcal{V}:u\in\mathcal{U}} \bar{\Psi}_{1,t}(\mathfrak{P}) \\ &= \bar{\Psi}_{1,t}(\mathfrak{P}_*) + o_t(1) \\ &= \Psi_{\mathrm{RS}} + o_t(1).\end{aligned}$$

This shows the upper bound of (36). Furthermore, by (173) and a Taylor expansion, for any $\mathfrak{P} \in \mathcal{V}$ with $u \in (0,K)$ and $|u-u_*| > \varsigma$, we have

$$\begin{aligned}\bar{\Psi}_{1,t}(\mathfrak{P}) &\le \bar{\Psi}_{1,t}(\mathfrak{P}_*) + \nabla\bar{\Psi}_{1,t}(\mathfrak{P}_*)^\top(\mathfrak{P}-\mathfrak{P}_*) \\ &\quad - \frac{1}{2}\mathfrak{e}^{1/2}\|\mathfrak{P}-\mathfrak{P}_*\|^2 \\ &\le \Psi_{\mathrm{RS}} + o_t(1)\cdot\|\mathfrak{P}-\mathfrak{P}_*\| - \frac{1}{2}\mathfrak{e}^{1/2}\varsigma^2.\end{aligned}$$

Applying the bound $\|\mathfrak{P}-\mathfrak{P}_*\| < C$ for a constant $C = C(K) > 0$ independent of t, and taking the limit $t \to \infty$, we obtain (37). ∎

Appendix D
Analysis of the Conditional Second Moment

In this appendix we prove Lemmas 3 and 4. We will abbreviate parts of the arguments that are similar to the preceding analysis of the conditional first moment, and also refer to [33, Lemmas 4.1 and 4.2] for some of the technical details.

Lemma 11: Let π be any probability distribution over $\mathbb{R}$. For $a, b \in \mathbb{R}^2$ and $c \in \mathbb{R}$, let $c_\pi(a,b,c)$ be as defined in (42), and let

$$\mathcal{O} = \left\{(a,c) \in \mathbb{R}^3 : \int e^{a_1 x_1^2 + a_2 x_2^2 + c x_1 x_2} d\pi(x_1) d\pi(x_2) < \infty\right\}.$$

Then $\mathcal{O}$ is a non-empty convex subset of $\mathbb{R}^3$. For any (a,c) in the interior of the complement of $\mathcal{O}$ and any $b \in \mathbb{R}^2$, we have $c_\pi(a,b,c) = \infty$. For any (a,c) in the interior of $\mathcal{O}$, the function $b \mapsto \log c_\pi(a,b,c)$ is continuous and satisfies, for some (a,c,π)-dependent constant $C > 0$ and for all $b \in \mathbb{R}^2$,

$$\log c_\pi(a,b,c) \le C(\|b\|^2 + 1).$$

Proof: The set $\mathcal{O}$ is convex by convexity of the function $(a,c) \mapsto c_\pi(a,0,c)$, and non-empty because $(a,c) = 0$ belongs to $\mathcal{O}$. The proofs of the remaining statements are similar to the proof of Lemma 9 and omitted for brevity. ∎

Proof of Lemma 3: Fix t and write $\mathcal{G}, X, Y, S, \Delta$ for $\mathcal{G}_t, X_t, Y_t, S_t, \Delta_t$. Then

$$\mathbb{E}[\mathcal{Z}(\overline{\mathcal{U}})^2 \mid \mathcal{G}] = \int \mathbb{I}\Big(\frac{1}{n}\|\tilde{\sigma}\|^2 \in \overline{\mathcal{U}},\ \frac{1}{n}\|\tilde{\tau}\|^2 \in \overline{\mathcal{U}}\Big) \times \exp\Big(-\|\epsilon\|^2 + \frac{n}{2}\cdot f_n(\tilde{\sigma},\tilde{\tau})\Big)\prod_{i=1}^n d\pi(\sigma_i)d\pi(\tau_i)$$

with $\tilde{\sigma} := \tilde{\sigma}(\sigma) = \sigma - \beta^\star$, $\tilde{\tau} := \tilde{\tau}(\tau) = \tau - \beta^\star$, and

$$f_n(\tilde{\sigma},\tilde{\tau}) := \frac{2}{n}\log\mathbb{E}\Big[\exp\Big(-\frac{\tilde{\sigma}^\top O^\top D^\top D O\tilde{\sigma}}{2} - \frac{\tilde{\tau}^\top O^\top D^\top D O\tilde{\tau}}{2} + (\tilde{\sigma}+\tilde{\tau})^\top O^\top D^\top \xi\Big)\,\Big|\,\mathcal{G}\Big].$$

Uniform Approximation of $f_n(\tilde{\sigma},\tilde{\tau})$: Define $\mathfrak{P}(\tilde{\sigma},\tilde{\tau}) = (u(\tilde{\sigma},\tilde{\tau}), r(\tilde{\sigma},\tilde{\tau}), v(\tilde{\sigma},\tilde{\tau}), w(\tilde{\sigma},\tilde{\tau}), p(\tilde{\sigma},\tilde{\tau}))$ by

$$u(\tilde{\sigma},\tilde{\tau}) = \frac{1}{n}\Big(\|\tilde{\tau}\|^2, \|\tilde{\sigma}\|^2\Big) \in \mathbb{R}^2, \qquad p(\tilde{\sigma},\tilde{\tau}) = \frac{1}{n}\tilde{\sigma}^\top\tilde{\tau} \in \mathbb{R},$$

$$\begin{pmatrix} r(\tilde{\sigma},\tilde{\tau})^\top \\ v(\tilde{\sigma},\tilde{\tau}) \\ w(\tilde{\sigma},\tilde{\tau})\end{pmatrix} = \left[\frac{1}{n}\begin{pmatrix} e^\top e & e^\top X & e^\top Y \\ X^\top e & X^\top X & X^\top Y \\ Y^\top e & Y^\top X & Y^\top Y\end{pmatrix}\right]^{-1/2} \times \frac{1}{n}(e,X,Y)^\top(\tilde{\sigma},\tilde{\tau}) \in \mathbb{R}^{(2t+1)\times 2}$$

where $r(\tilde{\sigma},\tilde{\tau}) \in \mathbb{R}^2$ and $v(\tilde{\sigma},\tilde{\tau}), w(\tilde{\sigma},\tilde{\tau}) \in \mathbb{R}^{t\times 2}$. Define the open domain

$$\mathcal{K} = \big\{\tilde{\sigma},\tilde{\tau} \in \mathbb{R}^n : u(\tilde{\sigma},\tilde{\tau}) \in (0,K)^2,\ \mathcal{A}(\mathfrak{P}(\tilde{\sigma},\tilde{\tau})) \succ 0\big\} \tag{174}$$

and write again $a_n(\tilde{\sigma},\tilde{\tau}) \doteq b_n(\tilde{\sigma},\tilde{\tau})$ to mean $a_n(\tilde{\sigma},\tilde{\tau}) - b_n(\tilde{\sigma},\tilde{\tau}) \to 0$ uniformly over $(\tilde{\sigma},\tilde{\tau}) \in \mathcal{K}$, almost surely as $n, m \to \infty$.

Recall $\tilde{\sigma}_\perp, \tilde{\sigma}_\parallel$ from (84), $\Pi = \Pi_{(e_b,S,\Lambda S)^\perp} \in \mathbb{R}^{n\times(n-2t-1)}$, and denote similarly

$$\tilde{\tau}_\perp = \Pi_{(e,X,Y)^\perp}^\top \tilde{\tau} \in \mathbb{R}^{n-2t-1},$$

$$\tilde{\tau}_\parallel = (e_b, S, \Lambda S)\begin{pmatrix} e^\top e & e^\top X & e^\top Y \\ X^\top e & X^\top X & X^\top Y \\ Y^\top e & Y^\top X & Y^\top Y\end{pmatrix}^{-1}(e,X,Y)^\top\tilde{\tau} \in \mathbb{R}^n.$$

Observe that, similarly to (90), we have

$$\mathcal{A}(\mathfrak{P}(\tilde{\sigma},\tilde{\tau})) = \frac{1}{n}\begin{pmatrix}\|\tilde{\sigma}_\perp\|^2 & \tilde{\sigma}_\perp^\top\tilde{\tau}_\perp \\ \tilde{\sigma}_\perp^\top\tilde{\tau}_\perp & \|\tilde{\tau}_\perp\|^2\end{pmatrix}.$$

The condition $\mathcal{A}(\mathfrak{P}(\tilde{\sigma},\tilde{\tau})) \succ 0$ defining $\mathcal{K}$ then requires that $\tilde{\sigma}_\perp$ and $\tilde{\tau}_\perp$ are non-zero and linearly independent. Then for a sufficiently small constant $\mathfrak{d} > 0$, an application of [34, Lemma B.2] and [33, Proposition 2.8] yields

$$f_n(\tilde{\sigma}) \doteq -\frac{1}{n}\left(\tilde{\sigma}_\parallel^\top D^\top D\tilde{\sigma}_\parallel + \tilde{\tau}_\parallel^\top D^\top D\tilde{\tau}_\parallel\right) + \frac{2}{n}(\tilde{\sigma}_\parallel + \tilde{\tau}_\parallel)^\top D^\top \xi + E_n(\tilde{\sigma},\tilde{\tau}) \tag{175}$$

where

$$\begin{aligned} E_n(\tilde{\sigma},\tilde{\tau}) = \inf_{\mathfrak{Z}\succeq(-d+\mathfrak{d})I} \Bigg\{ & \frac{1}{n}\mathrm{Tr}\left[\mathfrak{Z}\cdot\begin{pmatrix}\|\tilde{\sigma}_\perp\|^2 & \tilde{\sigma}_\perp^\top\tilde{\tau}_\perp \\ \tilde{\sigma}_\perp^\top\tilde{\tau}_\perp & \|\tilde{\tau}_\perp\|^2\end{pmatrix}\right] \\ &+ \frac{1}{n}\begin{pmatrix}\Pi^\top D^\top\xi - \Pi^\top D^\top D\tilde{\sigma}_\parallel \\ \Pi^\top D^\top\xi - \Pi^\top D^\top D\tilde{\tau}_\parallel\end{pmatrix}^\top \left(\mathfrak{Z}\oplus\Pi^\top D^\top D\Pi\right)^{-1} \\ &\times \begin{pmatrix}\Pi^\top D^\top\xi - \Pi^\top D^\top D\tilde{\sigma}_\parallel \\ \Pi^\top D^\top\xi - \Pi^\top D^\top D\tilde{\tau}_\parallel\end{pmatrix} \\ &- \frac{1}{n}\log\det\left(\mathfrak{Z}\oplus\Pi^\top D^\top D\Pi\right) \\ &- \left(2 + \log\det\frac{1}{n}\begin{pmatrix}\|\tilde{\sigma}_\perp\|^2 & \tilde{\sigma}_\perp^\top\tilde{\tau}_\perp \\ \tilde{\sigma}_\perp^\top\tilde{\tau}_\perp & \|\tilde{\tau}_\perp\|^2\end{pmatrix}\right)\Bigg\} \end{aligned} \tag{176}$$

and we define

$$\mathfrak{Z}\oplus\Pi^\top D^\top D\Pi := \mathfrak{Z}\otimes I_{(n-2t-1)\times(n-2t-1)} + I_{2\times 2}\otimes\Pi^\top D^\top D\Pi.$$

Similarly to (97), (98), and (99) from Lemma 1, we have

$$\frac{1}{n}(\tilde{\sigma}_\parallel^\top D^\top D\tilde{\sigma}_\parallel + \tilde{\tau}_\parallel^\top D^\top D\tilde{\tau}_\parallel) \doteq \sum_{i=1}^2 \frac{e_* r_i(\tilde{\sigma},\tilde{\tau})^2}{b_*} \tag{177}$$

$$+ \mathrm{Tr}\begin{pmatrix} d_* & a_* \\ a_* & c_*\end{pmatrix}\left(v_i(\tilde{\sigma},\tilde{\tau}), \frac{w_i(\tilde{\sigma},\tilde{\tau})}{\sqrt{\kappa_*}}\right)^\top \tag{178}$$

$$\times\left(v_i(\tilde{\sigma},\tilde{\tau}), \frac{w_i(\tilde{\sigma},\tilde{\tau})}{\sqrt{\kappa_*}}\right), \tag{179}$$

$$\frac{2}{n}(\tilde{\sigma}_\parallel + \tilde{\tau}_\parallel)^\top D\xi \doteq \sum_{i=1}^2 \frac{2r_i(\tilde{\sigma},\tilde{\tau})}{\sqrt{b_*}}, \tag{180}$$

and

$$\frac{1}{n}\mathrm{Tr}\left[\mathfrak{Z}\cdot\begin{pmatrix}\|\tilde{\sigma}_\perp\|^2 & \tilde{\sigma}_\perp^\top\tilde{\tau}_\perp \\ \tilde{\sigma}_\perp^\top\tilde{\tau}_\perp & \|\tilde{\tau}_\perp\|^2\end{pmatrix}\right] \tag{181}$$

$$- \frac{1}{n}\log\det\left(\mathfrak{Z}\oplus\Pi^\top D^\top D\Pi\right) \tag{182}$$

$$- \left(2 + \log\det\frac{1}{n}\begin{pmatrix}\|\tilde{\sigma}_\perp\|^2 & \tilde{\sigma}_\perp^\top\tilde{\tau}_\perp \\ \tilde{\sigma}_\perp^\top\tilde{\tau}_\perp & \|\tilde{\tau}_\perp\|^2\end{pmatrix}\right)$$

$$\doteq \mathcal{H}(\mathfrak{Z}, \mathcal{A}(\mathfrak{P}(\tilde{\sigma},\tilde{\tau}))). \tag{183}$$

This last approximation (183) holds uniformly over $\mathfrak{Z}\succeq(-d_- + \mathfrak{d})I$, by the same argument as in the proof of [33, Lemma 4.2].

For the remaining second term of $E_n(\tilde{\sigma},\tilde{\tau})$, write the eigen-decompositions

$$\mathfrak{Z} = \begin{pmatrix} y_1 & y_2\end{pmatrix}\begin{pmatrix}\zeta_1 & 0 \\ 0 & \zeta_2\end{pmatrix}\begin{pmatrix} y_1^\top \\ y_2^\top\end{pmatrix},$$

$$\Pi^\top D^\top D\Pi = \Pi'^\top D'\Pi' = \Pi'^\top \mathrm{diag}\left(d'_1,\ldots,d'_{n-2t-1}\right)\Pi'$$

where ζ_1,ζ_2 and $d'_1,\ldots,d'_{n-2t-1}$ are the eigenvalues of $\mathfrak{Z}$ and $\Pi^\top D^\top D\Pi$ respectively, and $y_1, y_2\in\mathbb{R}^2$ and the rows of $\Pi'\in\mathbb{R}^{(n-2t-1)\times(n-2t-1)}$ are the eigenvectors. Then

$$(\mathfrak{Z}\oplus\Pi^\top D^\top D\Pi)^{-1} = \begin{pmatrix}\Pi' & 0 \\ 0 & \Pi'\end{pmatrix}^\top \times\begin{pmatrix}\mathfrak{Z}_{11}\cdot I + D' & \mathfrak{Z}_{12}\cdot I \\ \mathfrak{Z}_{12}\cdot I & \mathfrak{Z}_{22}\cdot I + D'\end{pmatrix}^{-1}\begin{pmatrix}\Pi' & 0 \\ 0 & \Pi'\end{pmatrix}$$

and we may compute the inverse on the right side by inverting separately the 2×2 blocks,

$$\begin{pmatrix}\mathfrak{Z}_{11} + d'_i & \mathfrak{Z}_{12} \\ \mathfrak{Z}_{12} & \mathfrak{Z}_{22} + d'_i\end{pmatrix}^{-1} = \frac{1}{\zeta_1 + d'_i} y_1 y_1^\top + \frac{1}{\zeta_2 + d'_i} y_2 y_2^\top.$$

Then for each $j,k\in\{1,2\}$, the (j,k) block of $(\mathfrak{Z}\oplus\Pi^\top D^\top D\Pi)^{-1}$ is

$$(\mathfrak{Z}\oplus\Pi^\top D^\top D\Pi)^{-1}_{jk} = y_{1j}y_{1k}\left(\zeta_1 I + \Pi^\top D^\top D\Pi\right)^{-1} + y_{2j}y_{2k}\left(\zeta_2 I + \Pi^\top D^\top D\Pi\right)^{-1}. \tag{184}$$

Recall υ^A, υ^B from (110), and define $\mathfrak{X}^e(\zeta) = \mathfrak{X}^e_{22}(\zeta) - \mathfrak{X}^e_{12}(\zeta)\mathfrak{X}^e_{11}(\zeta)^{-1}\mathfrak{X}^e_{12}(\zeta)^\top\in\mathbb{R}^{2\times 2}$ where

$$\mathfrak{X}^e_{22}(\zeta) = \mathbb{E}\frac{\mathsf{E}_b^2}{\zeta + \mathsf{D}^2}\begin{pmatrix}\upsilon^A(\mathsf{D}^2) \\ \upsilon^B(\mathsf{D}^2)\end{pmatrix}\begin{pmatrix}\upsilon^A(\mathsf{D}^2) \\ \upsilon^B(\mathsf{D}^2)\end{pmatrix}^\top,$$

$$\mathfrak{X}^e_{12}(\zeta) = \mathbb{E}\frac{\mathsf{E}_b^2}{\zeta + \mathsf{D}^2}\begin{pmatrix}\upsilon^A(\mathsf{D}^2) \\ \upsilon^B(\mathsf{D}^2)\end{pmatrix}, \quad \mathfrak{X}^e_{11}(\zeta) = \mathbb{E}\frac{\mathsf{E}_b^2}{\zeta + \mathsf{D}^2}.$$

Note in particular that $\mathcal{F}^e$ defined in (105) is given by $\mathcal{F}^e(\zeta, r) = (r\ \ 1)\mathfrak{X}^e(\zeta)(r\ \ 1)^\top$. Then, applying the argument leading to (122) separately for each of the four blocks $j,k\in\{1,2\}$ of (184), we have

$$\frac{1}{n}\begin{pmatrix}\Pi^\top D^\top\xi - \Pi^\top D^\top D\tilde{\sigma}_\parallel \\ \Pi^\top D^\top\xi - \Pi^\top D^\top D\tilde{\tau}_\parallel\end{pmatrix}^\top\left(\mathfrak{Z}\oplus\Pi^\top D^\top D\Pi\right)^{-1} \times\begin{pmatrix}\Pi^\top D^\top\xi - \Pi^\top D^\top D\tilde{\sigma}_\parallel \\ \Pi^\top D^\top\xi - \Pi^\top D^\top D\tilde{\tau}_\parallel\end{pmatrix} \tag{185}$$

$$\doteq \sum_{j,k=1}^2 y_{1j}y_{1k}\Bigg[\begin{pmatrix} r_j(\tilde{\sigma},\tilde{\tau}) & 1\end{pmatrix}\mathfrak{X}^e(\zeta_1)\begin{pmatrix} r_k(\tilde{\sigma},\tilde{\tau}) \\ 1\end{pmatrix} \tag{186}$$

$$+ \mathcal{F}(\zeta_1)\cdot\mathcal{B}(v(\tilde{\sigma},\tilde{\tau}), w(\tilde{\sigma},\tilde{\tau}))_{jk}\Bigg]$$

$$+ y_{2j}y_{2k}\Bigg[\begin{pmatrix} r_j(\tilde{\sigma},\tilde{\tau}) & 1\end{pmatrix}\mathfrak{X}^e(\zeta_2)\begin{pmatrix} r_k(\tilde{\sigma},\tilde{\tau}) \\ 1\end{pmatrix} \tag{187}$$

$$+ \mathcal{F}(\zeta_2)\cdot\mathcal{B}(v(\tilde{\sigma},\tilde{\tau}), w(\tilde{\sigma},\tilde{\tau}))_{jk}\Bigg]$$

$$= \mathrm{Tr}\Bigg[\sum_{i=1}^2\begin{pmatrix} r_1(\tilde{\sigma},\tilde{\tau}) & r_2(\tilde{\sigma},\tilde{\tau}) \\ 1 & 1\end{pmatrix}^\top\mathfrak{X}^e(\zeta_i) \tag{188}$$

$$\begin{pmatrix} r_1(\tilde{\sigma},\tilde{\tau}) & r_2(\tilde{\sigma},\tilde{\tau}) \\ 1 & 1\end{pmatrix} y_i y_i^\top\Bigg] \tag{189}$$

$$+ \mathrm{Tr}\Big[\mathcal{F}(\mathfrak{Z})\cdot\mathcal{B}(v(\tilde{\sigma},\tilde{\tau}), w(\tilde{\sigma},\tilde{\tau}))\Big] \tag{190}$$

uniformly over $\mathfrak{Z}\succeq(-d_- + \mathfrak{d})I$, where $\mathcal{F}$ is applied in the second term spectrally to $\mathfrak{Z}$ via functional calculus.

From the form of $\mathcal{F}$ in (113), the second term of (190) may be expressed as

$$\begin{aligned}&\mathrm{Tr}[\mathcal{F}(\mathfrak{Z})\cdot\mathcal{B}(v,w)]\\&=\inf_{\chi^B,\chi^C\in\mathbb{R}^2}\sum_{i=1}^{2}\mathbb{E}\Big[\frac{1}{\zeta_i+\mathsf{D}^2}\Big(\theta\left(\mathsf{D}^2\right)\\&\quad-\lambda\left(\mathsf{D}^2\right)\chi_i^B-\chi_i^C\Big)^2\Big]y_i^\top\mathcal{B}(v,w)y_i. \qquad (191)\end{aligned}$$

For the first term, for $r=(r_1,r_2)\in\mathbb{R}^2$, set

$$\mathfrak{f}^e(x,r)=\begin{pmatrix}\mathfrak{f}^e(x,r_1)\\\mathfrak{f}^e(x,r_2)\end{pmatrix}=\begin{pmatrix}r_1&r_2\\1&1\end{pmatrix}^\top\begin{pmatrix}\upsilon^A(x)\\\upsilon^B(x)\end{pmatrix}\in\mathbb{R}^2,$$

where $\mathfrak{f}^e(x,r_i)$ is the function defined in (110). Then define

$$\begin{aligned}\mathcal{F}^e(\mathfrak{Z},r)&=\inf_{\chi^A\in\mathbb{R}^2}\mathbb{E}\Big[\Big(\begin{pmatrix}\frac{\gamma_*}{\eta_*}-b_*^{-1/2}r_1\\\frac{\gamma_*}{\eta_*}-b_*^{-1/2}r_2\end{pmatrix}\mathsf{D}^2-\chi^A\Big)^\top \qquad (192)\\&\quad\times\left(\mathsf{E}_b^2(\mathfrak{Z}+\mathsf{D}^2\cdot I)^{-1}\right) \qquad (193)\\&\quad\times\Big(\begin{pmatrix}\frac{\gamma_*}{\eta_*}-b_*^{-1/2}r_1\\\frac{\gamma_*}{\eta_*}-b_*^{-1/2}r_2\end{pmatrix}\mathsf{D}^2-\chi^A\Big)\Big]\\&=\inf_{\chi^A\in\mathbb{R}^2}\mathbb{E}\Big[\left(\mathfrak{f}^e(\mathsf{D}^2,r)-\chi^A\right)^\top \qquad (194)\\&\quad\times\left(\mathsf{E}_b^2(\mathfrak{Z}+\mathsf{D}^2\cdot I)^{-1}\right)\left(\mathfrak{f}^e(\mathsf{D}^2,r)-\chi^A\right)\Big] \qquad (195)\end{aligned}$$

where these expressions are equivalent by an additive shift of χ^A. Evaluating explicitly the infimum over χ^A, we get $\mathcal{F}^e(\mathfrak{Z},r)=\mathcal{F}^e_{22}(\mathfrak{Z},r)-\mathcal{F}^e_{12}(\mathfrak{Z},r)^\top\mathcal{F}^e_{11}(\mathfrak{Z})^{-1}\mathcal{F}^e_{12}(\mathfrak{Z},r)$ where

$$\begin{aligned}\mathcal{F}^e_{22}(\mathfrak{Z},r)&=\mathbb{E}\,\mathfrak{f}^e(\mathsf{D}^2,r)^\top\left(\mathsf{E}_b^2(\mathfrak{Z}+\mathsf{D}^2\cdot I)^{-1}\right)\mathfrak{f}^e(\mathsf{D}^2,r)\\&=\mathbb{E}\sum_{i=1}^{2}\begin{pmatrix}\upsilon^A(\mathsf{D}^2)\\\upsilon^B(\mathsf{D}^2)\end{pmatrix}^\top\begin{pmatrix}r_1&r_2\\1&1\end{pmatrix}\cdot\frac{\mathsf{E}_b^2y_iy_i^\top}{\zeta_i+\mathsf{D}^2}\\&\quad\times\begin{pmatrix}r_1&r_2\\1&1\end{pmatrix}^\top\begin{pmatrix}\upsilon^A(\mathsf{D}^2)\\\upsilon^B(\mathsf{D}^2)\end{pmatrix}\\&=\mathrm{Tr}\left[\sum_{i=1}^{2}\begin{pmatrix}r_1&r_2\\1&1\end{pmatrix}^\top\mathfrak{X}^e_{22}(\zeta_i)\begin{pmatrix}r_1&r_2\\1&1\end{pmatrix}y_iy_i^\top\right],\\\mathcal{F}^e_{12}(\mathfrak{Z},r)&=\mathbb{E}\,\mathsf{E}_b^2(\mathfrak{Z}+\mathsf{D}^2\cdot I)^{-1}\mathfrak{f}^e(\mathsf{D}^2,r)\\&=\mathbb{E}\sum_{i=1}^{2}\frac{\mathsf{E}_b^2y_iy_i^\top}{\zeta_i+\mathsf{D}^2}\cdot\begin{pmatrix}r_1&r_2\\1&1\end{pmatrix}^\top\begin{pmatrix}\upsilon^A(\mathsf{D}^2)\\\upsilon^B(\mathsf{D}^2)\end{pmatrix}\\&=\sum_{i=1}^{2}y_iy_i^\top\begin{pmatrix}r_1&r_2\\1&1\end{pmatrix}^\top\mathfrak{X}^e_{12}(\zeta_i),\\\mathcal{F}^e_{11}(\mathfrak{Z})&=\mathbb{E}\,\mathsf{E}_b^2(\mathfrak{Z}+\mathsf{D}^2\cdot I)^{-1}=\mathbb{E}\sum_{i=1}^{2}\frac{\mathsf{E}_b^2y_iy_i^\top}{\zeta_i+\mathsf{D}^2}\\&=\sum_{i=1}^{2}y_iy_i^\top\mathfrak{X}^e_{11}(\zeta_i). \qquad (196)\end{aligned}$$

Then it follows that the first term of (190) has the form

$$\mathrm{Tr}\left[\sum_{i=1}^{2}\begin{pmatrix}r_1&r_2\\1&1\end{pmatrix}^\top\mathfrak{X}^e(\zeta_i)\begin{pmatrix}r_1&r_2\\1&1\end{pmatrix}y_iy_i^\top\right]=\mathcal{F}^e(\mathfrak{Z},r). \qquad (197)$$

Combining (175), (176), (179), (180), (183), (190), and (197), we obtain the uniform approximation

$$\lim_{n,m\to\infty}\sup_{(\tilde\sigma,\tilde\tau)\in\mathcal{K}}\left|f_n(\tilde\sigma,\tilde\tau)-f(\mathfrak{P}(\tilde\sigma,\tilde\tau))\right|=0$$

where we define on the domain $\mathcal{V}=\{\mathfrak{P}:\mathcal{A}(\mathfrak{P})\succ0\}$ the function

$$\begin{aligned}f(\mathfrak{P})=\inf_{\mathfrak{Z}}\Big(&-\frac{e_*\|r\|^2}{b_*}+\frac{2r^\top1_{2\times1}}{\sqrt{b_*}}\\&-\mathrm{Tr}\begin{pmatrix}d_*&a_*\\a_*&c_*\end{pmatrix}\Big[\Big(v_1,\frac{w_1}{\sqrt{\kappa_*}}\Big)^\top\Big(v_1,\frac{w_1}{\sqrt{\kappa_*}}\Big)\\&+\Big(v_2,\frac{w_2}{\sqrt{\kappa_*}}\Big)^\top\Big(v_2,\frac{w_2}{\sqrt{\kappa_*}}\Big)\Big]\\&+\mathcal{H}(\mathfrak{Z},\mathcal{A}(\mathfrak{P}))+\mathcal{F}^e(\mathfrak{Z},r)+\mathrm{Tr}[\mathcal{F}(\mathfrak{Z})\cdot\mathcal{B}(v,w)]\Big)\end{aligned}$$

and the infimum is over $\mathfrak{Z}\succeq(-d_-+\mathfrak{d})I$. It is immediate from the forms (191) and (195) that $\mathcal{F}^e(\mathfrak{Z},r)$ and $\mathrm{Tr}[\mathcal{F}(\mathfrak{Z})\cdot\mathcal{B}(v,w)]$ are decreasing in the eigenvalues ζ_1,ζ_2 of $\mathfrak{Z}$. The same argument as in the proof of [33, Lemma 4.2] shows that $\mathcal{H}(\mathfrak{Z},\mathcal{A}(\mathfrak{P}))$ is also decreasing in each eigenvalue ζ_1,ζ_2 over the range $(-d_-,-d_-+\mathfrak{d}]$, and hence this infimum may be extended to the domain $\mathfrak{Z}\succ-d_-\cdot I$. Finally, since f_n is continuous on $\overline{\mathcal{K}}$ and the map $\mathfrak{P}:\overline{\mathcal{K}}\to\{\mathfrak{P}\in\overline{\mathcal{V}}:u_1,u_2\in[0,K]\}$ is continuous, relatively open, and maps $\mathcal{K}$ to the interior $\{\mathfrak{P}\in\mathcal{V}:u_1,u_2\in(0,K)\}$ for each fixed n, [33, Proposition C.1] shows that f extends continuously to $\{\mathfrak{P}\in\overline{\mathcal{V}}:u_1,u_2\in[0,K]\}$, and

$$\lim_{n,m\to\infty}\sup_{(\tilde\sigma,\tilde\tau)\in\overline{\mathcal{K}}}\left|f_n(\tilde\sigma,\tilde\tau)-f(\mathfrak{P}(\tilde\sigma,\tilde\tau))\right|=0.$$

Then, writing $\langle\cdot\rangle_\pi$ for the expectation over $(\sigma_i)_{i=1}^n,(\tau_i)_{i=1}^n\overset{iid}{\sim}\pi$, we obtain

$$\begin{aligned}&\lim_{n,m\to\infty}\frac{1}{n}\log\mathbb{E}[\mathcal{Z}(\overline{\mathcal{U}})^2\mid\mathcal{G}]=\\&\quad-1+\lim_{n,m\to\infty}\frac{1}{n}\log\Big\langle\mathbb{I}\left\{u(\tilde\sigma,\tilde\tau)\in\overline{\mathcal{U}}\times\overline{\mathcal{U}}\right\}\\&\qquad\exp\Big(\frac{n}{2}f(\mathfrak{P}(\tilde\sigma,\tilde\tau))\Big)\Big\rangle_\pi.\end{aligned}$$

Large Deviations Analysis: Introduce dual variables $\mathfrak{R}=(U,R,V,W,P)$ where $U,R\in\mathbb{R}^2$, $V,W\in\mathbb{R}^{t\times2}$, and $P\in\mathbb{R}$. Define

$$\begin{aligned}\lambda_n(\mathfrak{R})&=\frac{1}{n}\log\left\langle\exp\left(n\cdot\mathfrak{P}(\tilde\sigma,\tilde\tau)^\top\mathfrak{R}\right)\right\rangle_\pi\\\lambda(\mathfrak{R})&=\mathbb{E}\log c_\pi(U,\mathsf{A}(\mathfrak{R}),P)+\rho_*U^\top1_{2\times1}\\&\quad-\pi_*V_1^\top\Delta_t^{-1/2}1_{t\times1}-\pi_*V_2^\top\Delta_t^{-1/2}1_{t\times1}+\rho_*P\end{aligned}$$

where c_π is the function from (42), and where

$$\begin{aligned}\mathsf{A}(\mathfrak{R})&=-2U\mathsf{X}^\star+\frac{R\,\mathsf{E}}{\sqrt{b_*}}+V^\top\Delta_t^{-1/2}(\mathsf{X}_1,\ldots,\mathsf{X}_t)\\&\quad+\frac{W^\top\Delta_t^{-1/2}(\mathsf{Y}_1,\ldots,\mathsf{Y}_t)}{\sqrt{\kappa_*}}-P\mathsf{X}^\star1_{2\times1}\in\mathbb{R}^2.\end{aligned}$$

Then, applying Lemma 11 and the same argument as leading to (125), we have almost surely

- $\lim_{n,m\to\infty}\lambda_n(\mathfrak{R}) = \lambda(\mathfrak{R}) < \infty$ if (U,P) is in the interior of $\mathcal{O}$
- $\lim_{n,m\to\infty}\lambda_n(\mathfrak{R}) = \lambda(\mathfrak{R}) = \infty$ if (U,P) is in the interior of the complement of $\mathcal{O}$

where $\mathcal{O}$ is the convex set defined in Lemma 11. Under the sub-Gaussian condition (5), $(U,P)=0$ belongs to the interior of $\mathcal{O}$, so the Fenchel-Legendre dual λ^* of λ is a good convex rate function [47, Lemma 2.3.9(a)]. Defining $\bar{\lambda}(\mathfrak{R}) = \limsup_{n,m\to\infty}\lambda_n(\mathfrak{R})$, this coincides with $\lambda(\mathfrak{R})$ whenever $(U,P)\notin\partial\mathcal{O}$, the boundary of $\mathcal{O}$. For $(U,P)\in\partial\mathcal{O}$, since $\mathcal{O}$ is convex and 0 belongs to the interior of $\mathcal{O}$, the open line segment $\{s\cdot(U,P) : s\in[0,1)\}$ also belongs to the interior of $\mathcal{O}$ [48, Theorem 6.1]. The Gale-Klee-Rockafellar Theorem shows that $\lambda,\bar{\lambda}$ are upper-semicontinuous on $\{s\cdot(U,P) : s\in[0,1]\}$, so the supremum defining the Fenchel-Legendre dual $\lambda^*(\mathfrak{P}) = \sup_{\mathfrak{R}}\mathfrak{P}^\top\mathfrak{R} - \lambda(\mathfrak{R})$ may then be restricted to $\mathfrak{R}$ where $(U,P)\notin\partial\mathcal{O}$, and similarly for $\bar{\lambda}$. Then λ^* coincides with the Fenchel-Legendre dual of $\bar{\lambda}$, and the proof is concluded as in Lemma 1 by an application of the upper bound in the Gärtner-Ellis Theorem and Varadhan's lemma. ■

We now prove Lemma 4. Let $e_t = (0,\ldots,0,1)\in\mathbb{R}^t$, and consider $\mathfrak{P}_*,\mathfrak{Q}_*$ with the components

$$\begin{aligned}
&u_* = \frac{2}{\eta_*}1_{2\times1}, \quad r_* = \frac{\gamma_*}{\eta_*}b_*^{1/2}1_{2\times1},\\
&v_* = \frac{\eta_*-\gamma_*}{\eta_*}\Delta_t^{1/2}(e_t,e_t),\\
&w_* = \frac{\gamma_*}{\eta_*}\kappa_*^{1/2}\Delta_t^{1/2}(e_t,e_t), \quad p_* = \frac{1}{\eta_*}\\
&\mathfrak{Z}_* = (\eta_*-\gamma_*)\cdot I_{2\times2}, \quad \chi_*^A = \chi_*^B = \chi_*^C = 0,\\
&U_* = -\frac{\gamma_*}{2}1_{2\times1}, \quad R_* = \gamma_* b_*^{1/2}1_{2\times1}, \quad V_* = 0,\\
&W_* = \gamma_*\kappa_*^{1/2}\Delta_t^{1/2}(e_t,e_t), \quad P_* = 0.
\end{aligned}$$

Here, each of the two coordinates/columns of $u_*,r_*,v_*,w_*,U_*,R_*,V_*,W_*$ coincides with our previous specialization (130) in the analysis of the conditional first moment.

Lemma 12: In the setting of Lemma 4, for all $t\geq1$ and each $\iota\in\{u,r,v,w,p,\mathfrak{Z},\chi^A,\chi^B,\chi^C,U,R,W,P\}$,

$$\Phi_{2,t}(\mathfrak{P}_*,\mathfrak{Q}_*) = 2\Psi_{\mathrm{RS}}, \quad \partial_\iota\Phi_{2,t}(\mathfrak{P}_*,\mathfrak{Q}_*) = 0,$$
$$\lim_{t\to\infty}\|\partial_V\Phi_{2,t}(\mathfrak{P}_*,\mathfrak{Q}_*)\| = 0. \tag{198}$$

Proof: At $P_*=0$, we have $\log c_\pi(a_1,a_1,b_1,b_2,0) = \log c_\pi(a_1,b_1)+\log c_\pi(a_2,b_2)$ where c_π on the right side is defined by (33). At $\mathfrak{P}_*$, we have also $\mathcal{A}(\mathfrak{P}_*) = \eta_*^{-1}\cdot I_{2\times2}$ by the same calculation as (138), and $\mathcal{B}(v_*,w_*)=0$ because $v_* = \alpha_*^A w_*$. Since both $\mathcal{A}(\mathfrak{P}_*)$ and $\mathfrak{Z}_*$ are diagonal, it is then easily checked that

$$\Phi_{2,t}(\mathfrak{P}_*,\mathfrak{Q}_*) = 2\Phi_{1,t}(\mathfrak{P}_*,\mathfrak{Q}_*)$$

where $\mathfrak{P}_*,\mathfrak{Q}_*$ on the right side are the specializations of (130) from our previous analysis of the conditional first moment. Then $\Phi_{2,t}(\mathfrak{P}_*,\mathfrak{Q}_*) = 2\Psi_{\mathrm{RS}}$ follows from Lemma 10.

To check the stationarity conditions, define

$$\begin{aligned}
\mathsf{A}_* &= -2U_*\mathsf{X}^\star + b_*^{-1/2}R_*\mathsf{E} + V_*^\top\Delta_t^{-1/2}(\mathsf{X}_1,\ldots,\mathsf{X}_t)\\
&\quad + \kappa_*^{-1/2}W_*^\top\Delta_t^{-1/2}(\mathsf{Y}_1,\ldots,\mathsf{Y}_t) - P_*\mathsf{X}^\star 1_{2\times1}\\
&= \gamma_*(\mathsf{X}^\star+\mathsf{E}+\mathsf{Y}_t)1_{2\times1}.
\end{aligned}$$

Then we obtain, similarly to (140), (141), and (144),

$$\partial_{\mathfrak{Z}}\mathcal{H}(\mathfrak{Z}_*,\mathcal{A}(\mathfrak{P}_*)) = 0, \quad \partial_{\mathcal{A}}\mathcal{H}(\mathfrak{Z}_*,\mathcal{A}(\mathfrak{P}_*)) = -\gamma_*\cdot I_{2\times2},$$

$$\partial_r,\partial_{\mathfrak{Z}},\partial_{\chi^A}\mathbb{E}\Bigg[\left(\left[\frac{\gamma_*}{\eta_*}1_{2\times1} - \frac{r_*}{\sqrt{b_*}}\right]\mathsf{D}^2 - \chi_*^A\right)^\top \times\left(\mathsf{E}_b^2(\mathfrak{Z}_*+\mathsf{D}^2\cdot I_{2\times2})^{-1}\right)\left(\left[\frac{\gamma_*}{\eta_*}1_{2\times1} - \frac{r_*}{\sqrt{b_*}}\right]\mathsf{D}^2 - \chi_*^A\right)\Bigg] = 0,$$

and

$$\begin{aligned}
\partial_{a_1}\log c_\pi(U_*,\mathsf{A}_*,P_*) &= \partial_{a_2}\log c_\pi(U_*,\mathsf{A}_*,P_*)\\
&= \gamma_*^{-1}f'(\mathsf{X}^\star+\mathsf{E}+\mathsf{Y}_t)\\
&\quad + f(\mathsf{X}^\star+\mathsf{E}+\mathsf{Y}_t)^2,\\
\partial_{b_1}\log c_\pi(U_*,\mathsf{A}_*,P_*) &= \partial_{b_2}\log c_\pi(U_*,\mathsf{A}_*,P_*)\\
&= f(\mathsf{X}^\star+\mathsf{E}+\mathsf{Y}_t),\\
\partial_c\log c_\pi(U_*,\mathsf{A}_*,P_*) &= f(\mathsf{X}^\star+\mathsf{E}+\mathsf{Y}_t)^2.
\end{aligned}$$

Using these, $\mathcal{B}(v_*,w_*),\partial_v\mathcal{B}(v_*,w_*),\partial_w\mathcal{B}(v_*,w_*) = 0$, and the identities (146) and (148), we have

$$\begin{aligned}
&\partial_{\mathfrak{Z}}\Phi_{2,t}(\mathfrak{P}_*,\mathfrak{Q}_*) = 0, \quad \partial_p\Phi_{2,t}(\mathfrak{P}_*,\mathfrak{Q}_*) = -P_* = 0,\\
&\partial_P\Phi_{2,t}(\mathfrak{P}_*,\mathfrak{Q}_*)\\
&= \mathbb{E}f(\mathsf{X}^\star+\mathsf{E}+\mathsf{Y}_t)^2 - 2\mathbb{E}\mathsf{X}^\star f(\mathsf{X}^\star+\mathsf{E}+\mathsf{Y}_t) - (p_*-\rho_*)\\
&= 0,
\end{aligned}$$

and the analyses of the remaining derivatives are the same as in Lemma 10. ■

Proof of Lemma 4: The proof is analogous to the upper bound of Lemma 2. We fix $\mathfrak{h}\in\mathbb{R}_+$ and specialize the dual variables $\mathfrak{Q}=\mathfrak{Q}(\mathfrak{P})$ as functions of $\mathfrak{P}=(u,r,v,w,p)$, given by

$$\begin{aligned}
&U(u) = U_* + \mathfrak{h}(u-u_*), \quad R(r) = R_* + \mathfrak{h}(r-r_*),\\
&V(v) = \mathfrak{h}(v-v_*), \quad W(w) = W_* + \mathfrak{h}(w-w_*),\\
&P(p) = \mathfrak{h}(p-p_*), \quad \mathfrak{Z} = G^{-1}(\mathcal{A}(\mathfrak{P})),\\
&\chi^A(r) = \left(\tfrac{\gamma_*(\eta_*-\gamma_*)}{\eta_*} - b_*^{-1}\right)1_{2\times1} + b_*^{-3/2}e_* r,\\
&\chi^B = \chi^C = 0
\end{aligned}\tag{199}$$

where $G^{-1}(\mathcal{A})$ is defined spectrally by functional calculus. Then

$$\begin{aligned}
\Psi_{2,t}(\mathfrak{P}) &= \inf_{\mathfrak{Q}}\Phi_{2,t}(\mathfrak{P},\mathfrak{Q}) \leq \Phi_{2,t}(\mathfrak{P},\mathfrak{Q}(\mathfrak{P}))\\
&=: \bar{\Psi}_{2,t}(\mathfrak{P}) = -1 + \mathrm{I} + \mathrm{II} + \frac{1}{2}\Big(\mathrm{III}+\mathrm{IV}+\mathrm{V}+\mathrm{VI}\Big)
\end{aligned}\tag{200}$$

where

$$\begin{aligned}
\mathrm{I} &= \mathbb{E}\log c_\pi\Big(U(u),\ -2\mathsf{X}^\star U(u) - \mathsf{X}^\star P(p)\,1_{2\times1}\\
&\qquad + \left(R(r)\quad V(v)^\top\quad W(w)^\top\right)^\top\mathsf{F},\ P(p)\Big)\\
\mathrm{II} &= -(u-\rho_*1_{2\times1})^\top U(u) - r^\top R(r)\\
&\quad - (v_1+\pi_*\Delta_t^{-1/2}1_{t\times1})^\top V_1(v)\\
&\quad - (v_2+\pi_*\Delta_t^{-1/2}1_{t\times1})^\top V_2(v)\\
&\quad - w_1^\top W_1(w) - w_2^\top W_2(w) - (p-\rho_*)P(p)
\end{aligned}$$

$$\mathrm{III} = -\frac{e_*\|r\|^2}{b_*} + \frac{2r^\top 1_{2\times 1}}{\sqrt{b_*}} - \mathrm{Tr}\begin{pmatrix} d_* & a_* \\ a_* & c_* \end{pmatrix} \times\left[\left(v_1, \frac{w_1}{\sqrt{\kappa_*}}\right)^\top \left(v_1, \frac{w_1}{\sqrt{\kappa_*}}\right) + \left(v_2, \frac{w_2}{\sqrt{\kappa_*}}\right)^\top \left(v_2, \frac{w_2}{\sqrt{\kappa_*}}\right)\right]$$
$$\mathrm{IV} = \mathcal{H}(\mathfrak{Z}(\mathfrak{P}), \mathcal{A}(\mathfrak{P})), \quad \mathrm{V} = \mathcal{F}^e_{22}(\mathfrak{Z}(\mathfrak{P}), r),$$
$$\mathrm{VI} = \mathrm{Tr}[\mathcal{F}_{22}(\mathfrak{Z}(\mathfrak{P})) \cdot \mathcal{B}(v, w)]. \tag{201}$$

Here F in I is the tuple of random variables defined in (152), the function $\mathcal{F}^e_{22}(\mathfrak{Z}, r) = \mathbb{E}[\mathfrak{f}^e(\mathsf{D}^2, r)^\top(\mathsf{E}_b^2(\mathfrak{Z} + \mathsf{D}^2 \cdot I)^{-1})\mathfrak{f}^e(\mathsf{D}^2, r)]$ in V is as previously defined in (196), and the function $\mathcal{F}_{22}(\zeta) = \mathbb{E}[\theta(\mathsf{D}^2)^2/(\zeta + \mathsf{D}^2)]$ in VI is as previously defined in (110) and applied to $\mathfrak{Z}$ by functional calculus. We observe that $\mathfrak{Q}(\mathfrak{P}_*) = \mathfrak{Q}_*$, so $\bar{\Psi}_{2,t}(\mathfrak{P}_*) = 2\Psi_{\mathrm{RS}}$. An analysis similar to that of Lemma 2 shows also that $\|\nabla\bar{\Psi}_{2,t}(\mathfrak{P}_*)\| = o_t(1)$.

We now show that for sufficiently small $\mathfrak{e}_0 = \mathfrak{e}_0(\mathfrak{C}, K) > 0$ and $\mathfrak{e} < \mathfrak{e}_0$, this function $\bar{\Psi}_{2,t}(\mathfrak{P})$ is concave over $\{\mathfrak{P} \in \mathcal{V} : u_1, u_2 \in (0, K)\}$, by analyzing the Hessian of each term I-VI. Fix $\mathfrak{P} = (u, r, v, w, p) \in \mathcal{V}$ with $u_1, u_2 \in (0, K)$, fix a unit vector $\mathfrak{P}' = (u', r', v', w', p')$, and define for $s > 0$

$$\mathfrak{P}(s) = (u(s), r(s), v(s), w(s), p(s)) = (u, r, v, w, p) + s \cdot (u', r', v', w', p').$$

For I, recalling the random vector $\mathsf{F} \in \mathbb{R}^{2t+1}$ from (152), define

$$\mathsf{Q}(x_1, x_2) = \begin{pmatrix} x_1^2 - 2x_1\mathsf{X}^\star \\ x_2^2 - 2x_2\mathsf{X}^\star \\ x_1x_2 - x_1\mathsf{X}^\star - x_2\mathsf{X}^\star \\ x_1\mathsf{F} \\ x_2\mathsf{F} \end{pmatrix} \in \mathbb{R}^{4t+5}.$$

Vectorize $\mathfrak{Q}(\mathfrak{P})$ in the corresponding order $(U_1, U_2, P, R_1, V_1, W_1, R_2, V_2, W_2) \in \mathbb{R}^{4t+5}$ and vectorize similarly $\mathfrak{P}'$, and denote

$$\langle f(x_1, x_2)\rangle_{\mathfrak{P}} = \frac{\int f(x_1, x_2) \exp\left(\mathfrak{Q}(\mathfrak{P})^\top \mathsf{Q}(x_1, x_2)\right) d\pi(x_1) d\pi(x_2)}{\int \exp\left(\mathfrak{Q}(\mathfrak{P})^\top \mathsf{Q}(x_1, x_2)\right) d\pi(x_1) d\pi(x_2)}.$$

Write $\mathbb{V}_{\mathfrak{P}}[\cdot]$ for the corresponding variance. Then we have, analogous to (159),

$$\partial_s^2 \mathrm{I} = \mathfrak{h}^2 \mathbb{E}[\mathbb{V}_{\mathfrak{P}}[\mathfrak{P}'^\top \mathsf{Q}(x_1, x_2)]].$$

Note that the distribution defining $\langle\cdot\rangle_{\mathfrak{P}}$ corresponds to μ in Assumption 1.3 with $k = 2$ and

$$\Gamma = \begin{pmatrix} U_1(u) & \frac{1}{2}P(p) \\ \frac{1}{2}P(p) & U_2(u) \end{pmatrix},$$
$$z = -2\mathsf{X}^\star U(u) - \mathsf{X}^\star P(p)\, 1_{2\times 1} + \left(R(r) \;\; V(v)^\top \;\; W(w)^\top\right)^\top \mathsf{F}.$$

By Proposition 8, we have $U_{1,*} = U_{2,*} = -\gamma_*/2 = -(d_*/2)(1 + O(\mathfrak{e}))$. We have also $P_* = 0$ and $u_1, u_2, p, u_{1,*}, u_{2,*}, p_* \in (0, K)$ under the conditions $2\eta_*^{-1} \in (0, K)$ and $\mathfrak{P} \in \mathcal{V}$. Thus, choosing $\mathfrak{h} < \mathfrak{h}_0$ for a sufficiently small constant $\mathfrak{h}_0$ depending only on $(K, \mathfrak{C})$, we have $\Gamma \prec (4\mathfrak{C})^{-1}I$, and also $\mathfrak{C}^{-1} - \gamma_{\max} \geq c(1 + d_*)$ and $-\gamma_{\min} < C(1 + d_\star)$ for its largest and smallest eigenvalues. The same arguments as leading to (161) show $\|z\|^2 \leq (\mathsf{X}^{\star 2} + (q^\top\mathsf{F})^2) \cdot O(1 + d_*)$, and hence by Assumption 1.3 and Proposition 16,

$$\partial_s^2 \mathrm{I}\big|_{s=0} = O(\mathfrak{h}^2).$$

The same arguments as in (162–163) show

$$\partial_s^2 \mathrm{II}\big|_{s=0} = -2\mathfrak{h}, \quad \partial_s^2 \mathrm{III}\big|_{s=0} = -2d_* \left\|(r', v', w')\right\|^2 + O(\mathfrak{e})$$

where $(r', v', w') \in \mathbb{R}^{4t+2}$ is its vectorization and $\|\cdot\|$ is its Euclidean norm. For IV, we have by [33, Proposition 2.9(b)] that $\mathrm{IV} = \mathrm{Tr} f(\mathcal{A}(\mathfrak{P}(s)))$ where $f(\alpha) = \int_0^\alpha R(z)dz$. For all sufficiently small $\mathfrak{e}$, we may integrate the series representation (219) for $R(z)$ term-by-term to write $f(\mathcal{A}(\mathfrak{P}(s)))$ as the convergent matrix series

$$f(\mathcal{A}(\mathfrak{P}(s))) = -\mathcal{A}(\mathfrak{P}(s)) + \sum_{k\geq 2} \frac{\kappa_k}{k} \mathcal{A}(\mathfrak{P}(s))^k$$

where $|\kappa_k| \leq \kappa_2(16\mathfrak{e})^{k-2}$ and $\kappa_2 = O(\mathfrak{e}^2)$. It is easily checked that at $s = 0$, we have $\|\partial_s^k \mathcal{A}(\mathfrak{P}(s))\| = O(1)$ for $k = 0, 1, 2$, and in particular $\partial_s^2 \mathcal{A}(\mathfrak{P}(s)) = -2(r'r'^\top + v'^\top v' + w'^\top w') \in \mathbb{R}^{2\times 2}$, with trace $-2\|(r', v', w')\|^2$. Then, differentiating $f(\mathcal{A}(\mathfrak{P}(s)))$ term-by-term and taking the trace, it follows that

$$\partial_s^2 \mathrm{IV}\big|_{s=0} = 2d_* \left\|(r', v', w')\right\|^2 + O(\mathfrak{e}^2).$$

For V, applying the series expansion (167) now to the matrix argument $z = \mathcal{A}(\mathfrak{P}(s))$ and differentiating term-by-term, we have

$$\sup_{x\in\mathrm{supp}(\mathsf{D}^2)} \left\|\partial_s^k(\mathfrak{Z}(\mathfrak{P}(s)) + xI)^{-1}\right\|\bigg|_{s=0} = O(1) \text{ for } k = 0, 1, 2. \tag{202}$$

Combining with the bound (169), we have as in the proof of Lemma 2 that $\partial_s^2 \mathrm{V}\big|_{s=0} = O(\mathfrak{e}^2)$. For VI, recalling the bound (171) and combining this with (202), we obtain at $s = 0$ that $\|\partial_s^k \mathcal{F}_{22}(\mathfrak{Z}(s))\| = O(\kappa_2\mathfrak{e}^2/\eta_*^2)$ for $k = 0, 1, 2$. As in the proof of Lemma 2, we have $\|\partial_s^k \mathcal{B}(v(s), w(s))\| = O(\max(1, \eta_*^2/\kappa_2))$ for $k = 0, 1, 2$. Then $\partial_s^2 \mathrm{VI}\big|_{s=0} = O(\mathfrak{e}^2)$.

Combining the above and setting $\mathfrak{h} = \mathfrak{e}^{1/2}$, we conclude that for $\mathfrak{e} < \mathfrak{e}_0(\mathfrak{C}, K)$,

$$\nabla^2\bar{\Psi}_{2,t}(\mathfrak{P}) \prec -\mathfrak{e}^{1/2} \cdot I \text{ for all } \mathfrak{P} \in \mathcal{V} \text{ with } u_1, u_2 \in (0, K).$$

Since $u_* \in \mathcal{U} \subseteq (0, K)$ by assumption, the same argument as in Lemma 2 shows $\sup_{\mathfrak{P}\in\mathcal{V}:u_1,u_2\in\mathcal{U}} \Psi_{2,t}(\mathfrak{P}) \leq \bar{\Psi}_{2,t}(\mathfrak{P}_*) + o_t(1) = 2\Psi_{\mathrm{RS}} + o_t(1)$, and taking the limit $t \to \infty$ concludes the proof. ■

APPENDIX E
CONCENTRATION OF THE LOG-PARTITION FUNCTION

Proof: [Proof of Lemma 5] Recall the representation (82) and (86) of the conditional law of $\mathcal{Z}(U)$ given $\mathcal{G}_t$,

$$\frac{1}{n}\log \mathcal{Z}(U)\bigg|_{\mathcal{G}_t} \overset{L}{=} \frac{1}{n}\log \int \prod_{i=1}^n d\pi(\sigma_i)\mathbb{I}\left(\frac{1}{n}\|\tilde{\sigma}\|^2 \in U\right) \times \exp\left(-\frac{\|\xi\|^2}{2} - \frac{(\Pi\tilde{O}\tilde{\sigma}_\perp + \tilde{\sigma}_\parallel)^\top D^\top D(\Pi\tilde{O}\tilde{\sigma}_\perp + \tilde{\sigma}_\parallel)}{2} + \left(\Pi\tilde{O}\tilde{\sigma}_\perp + \tilde{\sigma}_\parallel\right)^\top D^\top\xi\right) \tag{203}$$

where $\tilde{\sigma} = \sigma - \beta^\star$ and $\|\tilde{\sigma}_\parallel\|^2, \|\tilde{\sigma}_\perp\|^2 \leq \|\tilde{\sigma}\|^2$. We denote the right side of (203) as $F(\tilde{O})$, where $\tilde{O} \sim \mathrm{Haar}(\mathbb{SO}(n - (2t+1)))$ is independent of $\mathcal{G}_t$, and all other quantities defining

$F(\tilde{O})$ are $\mathcal{G}_t$-measurable. On the event $\mathcal{E}$, this integral in (203) is non-zero, so for any $\tilde{O} \in \mathbb{SO}(n-(2t+1))$ the Euclidean gradient of F is bounded as

$$\begin{aligned}\left\|\frac{\partial}{\partial \tilde{O}} F(\tilde{O})\right\|_F &\le \frac{1}{n} \sup_{\tilde{\sigma}: \frac{\|\tilde{\sigma}\|^2}{n} \in U} \Big\| -\tilde{\sigma}_\perp \tilde{\sigma}_\perp^\top \tilde{O}^\top \Pi^\top D^\top D \Pi \\ &\qquad - \tilde{\sigma}_\perp \tilde{\sigma}_\parallel^\top D^\top D \Pi + \tilde{\sigma}_\perp \xi^\top D \Pi \Big\|_F \\ &\le \frac{1}{n} \sup_{\tilde{\sigma}: \frac{\|\tilde{\sigma}\|^2}{n} \in U} \Big(\left\|D^\top D\right\|_{\text{op}} \|\tilde{\sigma}_\perp\|^2 \\ &\quad + \left\|D^\top D\right\|_{\text{op}} \|\tilde{\sigma}_\perp\| \left\|\tilde{\sigma}_\parallel\right\| + \|\tilde{\sigma}_\perp\| \left\|D^\top \xi\right\| \Big) \\ &\le 2K \|D^\top D\|_{\text{op}} + \sqrt{KL}\end{aligned}$$

where we applied $\|\tilde{O}\|_{\text{op}}, \|\Pi\|_{\text{op}} \le 1$, $\|D^\top \xi\|^2 \le Ln$, $\left\|uv^\top A\right\| \le \|A\|_{\text{op}} \left\|uv^\top\right\|_F = \|A\|_{\text{op}} \|u\| \|v\|$, and the bounds $\|\tilde{\sigma}_\perp\|, \left\|\tilde{\sigma}_\parallel\right\| \le \|\tilde{\sigma}\| \le \sqrt{nK}$ when $U \subseteq [0, K]$. Then, applying the assumption $\left\|D^\top D\right\|_{\text{op}} \to d_+$, this bound is less than $2K(d_+ + 1) + \sqrt{KL}$ for all large n. Then

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{n}\log \mathcal{Z}(\mathcal{U}) - \mathbb{E}\left[\frac{1}{n}\log \mathcal{Z}(\mathcal{U}) \,\middle|\, \mathcal{G}_t\right]\right| \ge \delta \,\middle|\, \mathcal{G}_t\right) \mathbb{I}\{\mathcal{E}\} \\ \le 2\exp\left(-\frac{(n-(2t+1)-2)\delta^2}{8(2K(d_+ + 1) + \sqrt{KL})^2}\right)\end{aligned}$$

by Gromov's inequality, see e.g. [49, Theorem 4.4.27]. Choosing $C(K, L, d_+) > 8(2K(d_+ + 1) + \sqrt{KL})^2$ strictly, the statement (47) thus holds for all sufficiently large n. ■

Proof of Corollary 3: Set $\overline{\mathcal{U}} = [0, K]$. Fix any $\delta' > 0$, and fix $L > 0$ such that the condition $\|D^\top \xi\|^2 \le Ln$ in Lemma 5 holds almost surely for all large n. By Lemma 5, for a constant $c_0 = c_0(\delta', K, L, \mathfrak{e}) > 0$, any $t \ge 1$, and all large n,

$$\mathbb{P}\left(\left|\frac{1}{n}\log \mathcal{Z}(\overline{\mathcal{U}}) - \mathbb{E}\left[\frac{1}{n}\log \mathcal{Z}(\overline{\mathcal{U}}) \,\middle|\, \mathcal{G}_t\right]\right| \ge \delta' \,\middle|\, \mathcal{G}_t\right) \mathbb{I}(\mathcal{E}) \le e^{-c_0 n}. \tag{204}$$

Now fix any $\delta \in (0, c_0/6)$. By Lemmas 1, 2, 3, and 4, for a large enough iteration $t = t(\delta) \ge 1$ and large enough $M = M(\delta) > 0$, almost surely

$$\limsup_{n,m\to\infty} \frac{1}{n} \log \mathbb{E}\left[\mathcal{Z}(\overline{\mathcal{U}}) \mid \mathcal{G}_t\right] \le \sup_{u \in \mathcal{U}} \Psi_{1,t}(u) < \Psi_{\text{RS}} + \delta \tag{205}$$

$$\liminf_{n,m\to\infty} \frac{1}{n} \log \mathbb{E}\left[\mathcal{Z}(\overline{\mathcal{U}}) \mid \mathcal{G}_t\right] \ge \sup_{u \in \mathcal{U}} \Psi_{1,t}^M(u) > \Psi_{\text{RS}} - \delta \tag{206}$$

$$\lim_{n,m\to\infty} \frac{1}{n} \log \mathbb{E}\left[\mathcal{Z}(\overline{\mathcal{U}})^2 \mid \mathcal{G}_t\right] \le \sup_{u \in \mathcal{U}} \Psi_{2,t}(u) < 2\Psi_{\text{RS}} + \delta. \tag{207}$$

Letting $\mathcal{E}$ be the ($\mathcal{G}_t$-measurable) event in Lemma 5, if the first condition of $\mathcal{E}$ does not hold, then we have $\mathbb{P}[\mathcal{Z}(\overline{\mathcal{U}}) = 0 \mid \mathcal{G}_t] = 1$ and hence $\log \mathbb{E}[\mathcal{Z}(\overline{\mathcal{U}}) \mid \mathcal{G}_t] = -\infty$. Thus the finite lower bound in (206) and the above choice of $L > 0$ imply that $\mathcal{E}$ holds almost surely for all large n. Then taking the expectation of (204) and applying the Borel-Cantelli lemma, this implies almost surely for all large n,

$$\frac{1}{n} \log \mathcal{Z}(\overline{\mathcal{U}}) < \mathbb{E}\left[\frac{1}{n} \log \mathcal{Z}(\overline{\mathcal{U}}) \,\middle|\, \mathcal{G}_t\right] + \delta'.$$

Then applying Jensen's inequality and (205), almost surely for all large n,

$$\frac{1}{n} \log \mathcal{Z}(\overline{U}) < \frac{1}{n} \log \mathbb{E}[\mathcal{Z}(\overline{\mathcal{U}}) \mid \mathcal{G}_t] + \delta' < \Psi_{\text{RS}} + \delta + \delta'. \tag{208}$$

For the complementary lower bound, let $\mathcal{E}'$ be the ($\mathcal{G}_t$-measurable) event where

$$\begin{aligned}\frac{1}{n} \log \frac{\mathbb{E}\left[\mathcal{Z}(\overline{\mathcal{U}}) \mid \mathcal{G}_t\right]}{2} &> \Psi_{\text{RS}} - \delta, \\ \frac{1}{n} \log \mathbb{E}\left[\mathcal{Z}(\overline{\mathcal{U}})^2 \mid \mathcal{G}_t\right] &< 2\Psi_{\text{RS}} + \delta.\end{aligned}$$

Then (206) and (207) show that $\mathcal{E}'$ holds almost surely for all large n. On $\mathcal{E} \cap \mathcal{E}'$,

$$\begin{aligned}&\mathbb{P}\left[\frac{1}{n} \log \mathcal{Z}(\overline{\mathcal{U}}) > \Psi_{\text{RS}} - \delta \,\middle|\, \mathcal{G}_t\right] \mathbb{I}(\mathcal{E} \cap \mathcal{E}') \\ &\overset{(a)}{\ge} \mathbb{P}\left[\frac{1}{n} \log \mathcal{Z}(\overline{\mathcal{U}}) \ge \frac{1}{n} \log \frac{\mathbb{E}\left[\mathcal{Z}(\overline{\mathcal{U}}) \mid \mathcal{G}_t\right]}{2} \,\middle|\, \mathcal{G}_t\right] \mathbb{I}(\mathcal{E} \cap \mathcal{E}') \\ &= \mathbb{P}\left[\mathcal{Z}(\overline{\mathcal{U}}) \ge \frac{\mathbb{E}\left[\mathcal{Z}(\overline{\mathcal{U}}) \mid \mathcal{G}_t\right]}{2} \,\middle|\, \mathcal{G}_t\right] \mathbb{I}(\mathcal{E} \cap \mathcal{E}') \\ &\overset{(b)}{\ge} \frac{\mathbb{E}\left[\mathcal{Z}(\overline{\mathcal{U}}) \mid \mathcal{G}_t\right]^2}{4\mathbb{E}\left[\mathcal{Z}(\overline{\mathcal{U}})^2 \mid \mathcal{G}_t\right]} \cdot \mathbb{I}(\mathcal{E} \cap \mathcal{E}') \overset{(c)}{>} e^{-3n\delta} \cdot \mathbb{I}(\mathcal{E} \cap \mathcal{E}')\end{aligned}$$

where (a) and (c) apply the definition of the event $\mathcal{E}'$, and (b) applies the Paley-Zygmund inequality. By our choice $\delta < c_0/6$ where c_0 is the constant in (204), on the event $\mathcal{E} \cap \mathcal{E}'$ this last quantity is bounded below by $e^{-c_0 n/2}$. This and (204) together imply that on the event $\mathcal{E} \cap \mathcal{E}'$,

$$\mathbb{E}\left[\frac{1}{n} \log \mathcal{Z}(\overline{\mathcal{U}}) \,\middle|\, \mathcal{G}_t\right] > \Psi_{\text{RS}} - \delta - \delta'.$$

Then multiplying (204) by $\mathbb{I}(\mathcal{E}')$, taking the expectation on both sides, and applying the Borel-Cantelli lemma and the statement that $\mathcal{E} \cap \mathcal{E}'$ holds almost surely for all large n, we get

$$\frac{1}{n} \log \mathcal{Z}(\overline{\mathcal{U}}) > \mathbb{E}\left[\frac{1}{n} \log \mathcal{Z}(\overline{\mathcal{U}}) \,\middle|\, \mathcal{G}_t\right] - \delta' > \Psi_{\text{RS}} - \delta - 2\delta' \tag{209}$$

almost surely for all large n. The result follows upon taking $\delta, \delta' \to 0$ in (208) and (209). ■

Appendix F
Proofs for Unbounded Support

In this appendix, we complete the proofs of Theorems 1.7, 1.9, and 1.10 in the more general setting of Assumption 1.3 where π may have unbounded support.

Proof of Theorem 1.7, Unbounded Support: We apply a truncation argument. First note from (30) that

$$\begin{aligned}\Psi_{\mathrm{RS}} &\overset{(a)}{\geq} -\frac{1}{2} - \frac{\gamma_* \rho_*}{2} + \frac{\gamma_*}{2\eta_*} - \frac{d_*}{2\eta_*} \\ &\quad + \mathbb{E} \log c_\pi \left(-\frac{1}{2}\gamma_*, \gamma_* \mathsf{X}^\star + \sqrt{\gamma_*}\mathsf{Z} \right) \\ &\overset{(b)}{\geq} -\frac{1}{2} - \gamma_* \rho_* + \frac{\gamma_*}{2\eta_*} - \frac{d_*}{2\eta_*}\end{aligned}$$

where in (a) we used that $R(z)$ is increasing by Lemma 13, so $R(z) \geq \lim_{z\to 0} R(z) = \mathbb{E}[-\mathsf{D}^2] = -d_*$, and in (b) we used Jensen's inequality and the condition that π has mean 0 to bound $\mathbb{E}\log c_\pi\left(-\frac{1}{2}\gamma_*, \gamma_*\mathsf{X}^\star + \sqrt{\gamma_*}\mathsf{Z}\right) \geq \int (-\frac{1}{2}\gamma_*) x^2 \, d\pi(x) = -\frac{1}{2}\gamma_*\rho_*$. Then applying $\eta_*^{-1} \leq \rho_* \leq \mathfrak{C}$ by Proposition 1, and $\gamma_* \in [d_*/2, 2d_*]$ for all $\mathfrak{e} < \mathfrak{e}_0(\mathfrak{C})$ sufficiently small by Proposition 8,

$$\Psi_{\mathrm{RS}} > -\frac{1}{2} - \frac{5d_*}{2}\mathfrak{C} =: \Psi_{\mathrm{LB}}. \tag{210}$$

Fix a constant $K > 6\mathfrak{C}$, let $\mathcal{U} = (0, K)$, and consider

$$\begin{aligned}\mathcal{Z} - \mathcal{Z}(\overline{\mathcal{U}}) &= \mathcal{Z}((K,\infty)) \\ &= \int \mathbb{I}\left(\frac{1}{n}\|\sigma - \beta^\star\|^2 > K\right) \\ &\quad \times \exp\left(-\frac{\|A\beta^\star + \epsilon - A\sigma\|^2}{2}\right)\prod_{i=1}^n d\pi(\sigma_i).\end{aligned}$$

Define the event $\mathcal{A}$ where

$$\begin{aligned}n^{-1}\|\beta^\star\|^2 < 2\mathfrak{C}, \quad n^{-1}\|Q\epsilon\|^2 < 2, \\ \min(\mathrm{diag}(D^\top D)) > d_* - 2\mathfrak{e}.\end{aligned}$$

Observe that under the conditions $\min(\mathrm{diag}(D^\top D)) \to d_- \geq d_* - \mathfrak{e}$ by (4) and Assumption 1.4, $\rho_* \leq \mathfrak{C}$ by (5), and the concentration bound of Proposition 15, this event $\mathcal{A}$ holds almost surely for all large n. On the event $\mathcal{A}$ and for σ satisfying $n^{-1}\|\sigma - \beta^\star\|^2 > K$, let us first bound

$$\begin{aligned}e^{-\frac{\|A\beta^\star + \epsilon - A\sigma\|^2}{2}} &= e^{-\frac{\|DO(\beta^\star - \sigma) + Q\epsilon\|^2}{2}} \\ &\leq e^{-\frac{\|DO(\beta^\star - \sigma)\|^2}{4} + \frac{\|Q\epsilon\|^2}{2}} \\ &\leq e^{\left(-\frac{(d_* - 2\mathfrak{e})}{4}K + 1\right)n}\end{aligned}$$

Note that this quantity is also bounded above trivially by $e^0 = 1$. Let us then apply $\|\sigma - \beta^\star\|^2 \leq 2\|\sigma\|^2 + 2\|\beta^\star\|^2$ to bound

$$\mathbb{I}(\mathcal{A}) \cdot \mathcal{Z}((K,\infty)) \leq \exp\left(\min\left(0, -\frac{d_* - 2\mathfrak{e}}{4}K + 1\right)n\right) \tag{211}$$

$$\begin{aligned}&\times \int \mathbb{I}\left(\frac{1}{n}\|\sigma\|^2 > \frac{K - 4\mathfrak{C}}{2}\right) \cdot \prod_{i=1}^n d\pi(\sigma_i) \\ &\leq 2\exp\Bigg(\min\left(0, -\frac{d_* - 2\mathfrak{e}}{4}K + 1\right)n \qquad (212) \\ &\quad - c_0 \min\left(\left(\frac{K - 6\mathfrak{C}}{2\mathfrak{C}}\right)^2, \left(\frac{K - 6\mathfrak{C}}{2\mathfrak{C}}\right)\right)n\Bigg) \qquad (213)\end{aligned}$$

where the second inequality applies Proposition 15 and that the mean of $n^{-1}\|\sigma\|^2$ is $\rho_* \leq \mathfrak{C}$ under π. If $d_* < 1$, then $\Psi_{\mathrm{LB}} > -(1/2)(1 + 5\mathfrak{C})$. Choosing $K > 6\mathfrak{C} + 2\mathfrak{C} \cdot \max(1, (2c_0)^{-1} \cdot (1 + 5\mathfrak{C}))$ and bounding the first term in the exponent of (213) by 0, we obtain almost surely

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathcal{Z}((K,\infty)) < \Psi_{\mathrm{LB}} < \Psi_{\mathrm{RS}}. \tag{214}$$

If $d_* \geq 1$, then assuming $\mathfrak{e} < \mathfrak{e}_0 < 1/4$, the first term in the exponent of (213) is at most $[-(d_*/8)K + 1]n$. Then choosing $K > 20\mathfrak{C} + 12$ and bounding the second term in the exponent of (213) by 0 again ensures (214). In either case, the choice of K depends only on $\mathfrak{C}$. Combining with $n^{-1}\log \mathcal{Z}(\overline{\mathcal{U}}) \to \Psi_{\mathrm{RS}}$ from Corollary 3, which holds for $\mathfrak{e} < \mathfrak{e}_0(K, \mathfrak{C})$ sufficiently small, this shows $n^{-1}\log \mathcal{Z} \to \Psi_{\mathrm{RS}}$ almost surely.

To apply the dominated convergence theorem, observe that the right side of (50) may be bounded using $\int n^{-1}\|\beta^\star - \sigma\|^2 \prod_i d\pi(\sigma_i) \leq (2/n)\|\beta^\star\|^2 + 2\rho_*$, and that $\{\|\beta^\star\|^2/n\}_{n\geq 1}$ is uniformly integrable by the tail bound of Proposition 15. Then the dominated convergence theorem yields $n^{-1}\mathbb{E}[\log \mathcal{Z} \mid A] \to \Psi_{\mathrm{RS}}$ almost surely, and the remainder of the proof is the same as in the setting where π has bounded support. ■

Proof of Theorem 1.9, Unbounded Support: Recall Ψ_{LB} from (210), and consider again $\mathcal{Z}((K,\infty))$ for $K > 6\mathfrak{C}$. Recall the bound (214), and note that this bound holds simultaneously for every $K > 6\mathfrak{C}$ on the event $\mathcal{A}$ which holds almost surely for all large n. Applying this bound with $K(t) = 6\mathfrak{C} + 2\mathfrak{C} \cdot \min(1, -(2c_0)^{-1} \cdot (1 + 5\mathfrak{C})) + t$ when $d_* < 1$ and with $K(t) = 20\mathfrak{C} + 12 + 4c_0 t/\mathfrak{C}$ when $d_* \geq 1$ and $\mathfrak{e} < \mathfrak{e}_0 < 1/4$ gives

$$\mathbb{I}(\mathcal{A}) \cdot \mathcal{Z}((K(t),\infty)) \leq 2\exp\left(\Psi_{\mathrm{RS}} \cdot n - \frac{c_0 t}{2\mathfrak{C}} \cdot n\right).$$

Write as shorthand $X(\sigma, \beta^\star) = n^{-1}\|\sigma - \beta^\star\|^2$. Applying $\mathbb{E}[X \cdot \mathbb{I}(X > t)] = \int_0^\infty \mathbb{P}[X > \max(s,t)]ds$ for any nonnegative random variable X, we then have

$$\begin{aligned}&\mathbb{I}(\mathcal{A})\Big\langle X(\sigma,\beta^\star) \cdot \mathbb{I}\Big(X(\sigma,\beta^\star) > K(1)\Big)\Big\rangle \\ &= \int_0^\infty \mathbb{I}(\mathcal{A}) \cdot \Big\langle \mathbb{I}\Big(X(\sigma,\beta^\star) > \max(s, K(1))\Big)\Big\rangle ds \\ &= \frac{1}{\mathcal{Z}}\int_0^\infty \mathbb{I}(\mathcal{A}) \cdot \mathcal{Z}\Big((\max(s, K(1)), \infty)\Big) ds \\ &\leq \frac{2C}{\mathcal{Z}}\exp\left(\Psi_{\mathrm{RS}} \cdot n - \frac{c_0}{2\mathfrak{C}} \cdot n\right)\end{aligned}$$

for a constant $C > 0$ depending only on $(K(1), \mathfrak{C}, c_0)$. The event $\mathcal{A}$ holds almost surely for all large n, and the preceding proof of Theorem 1.7 verifies $n^{-1}\log \mathcal{Z} \to \Psi_{\mathrm{RS}}$ almost surely. Writing as shorthand $K = K(1)$, this shows that almost surely

$$\lim_{n,m\to\infty} \Big\langle X(\sigma,\beta^\star) \cdot \mathbb{I}\Big(X(\sigma,\beta^\star) > K\Big)\Big\rangle = 0. \tag{215}$$

Fixing any small constant $\varsigma > 0$ and defining $\mathcal{U} = (0, K) \setminus (2\eta_*^{-1} - \varsigma, 2\eta_*^{-1} + \varsigma)$, the proof in the setting of bounded support shows

$$\lim_{n,m\to\infty} \Big\langle \mathbb{I}\Big(X(\sigma,\beta^\star) \in \overline{\mathcal{U}}\Big)\Big\rangle = 0 \tag{216}$$

where $\overline{\mathcal{U}}$ is the closure of $\mathcal{U}$. Then, applying $2\eta_*^{-1} \le 2\rho_* \le 2\mathfrak{C}$,

$$\begin{aligned}&\left\langle\left|X(\sigma,\beta^\star)-2\eta_*^{-1}\right|\right\rangle\\ &\le \varsigma + \left\langle\left|X(\sigma,\beta^\star)-2\eta_*^{-1}\right|\cdot\mathbb{I}\Big(X(\sigma,\beta^\star)\in\overline{\mathcal{U}}\Big)\right\rangle\\ &\quad+\left\langle\left|X(\sigma,\beta^\star)-2\eta_*^{-1}\right|\cdot\mathbb{I}\Big(X(\sigma,\beta^\star)>K\Big)\right\rangle\\ &\le \varsigma+(K+2\mathfrak{C})\cdot\left\langle\mathbb{I}\Big(X(\sigma,\beta^\star)\in\overline{\mathcal{U}}\Big)\right\rangle\\ &\quad+\left\langle\Big(X(\sigma,\beta^\star)+2\mathfrak{C}\Big)\cdot\mathbb{I}\Big(X(\sigma,\beta^\star)>K\Big)\right\rangle.\end{aligned}$$

This last bound is at most 2ς almost surely for all large n by (215) and (216). Thus, almost surely

$$\lim_{n,m\to\infty}\frac{1}{2n}\langle\|\sigma-\beta^\star\|^2\rangle=\lim_{n,m\to\infty}\frac{X(\sigma,\beta^\star)}{2}=\eta_*^{-1}.$$

To apply the dominated convergence theorem, note that $(2n)^{-1}\langle\|\sigma-\beta^\star\|^2\rangle \le \langle\|\sigma\|^2/n\rangle+\|\beta^\star\|^2/n$. Here $\{\|\beta^\star\|^2/n\}_{n\ge1}$ is uniformly integrable by the tail bound of Proposition 15, and $\{\langle\|\sigma\|^2/n\rangle\}_{n\ge1}$ is uniformly integrable as it is uniformly bounded in L^2:

$$\begin{aligned}&\mathbb{E}\left[\left\langle\frac{\|\sigma\|^2}{n}\right\rangle^2\right]\le\mathbb{E}\left[\left\langle\frac{\|\sigma\|^4}{n^2}\right\rangle\right]\\ &=\mathbb{E}\left[\frac{\|\sigma\|^4}{n^2}\right]\le\frac{1}{n}\sum_{i=1}^n\mathbb{E}[\sigma_i^4]\\ &=\mathbb{E}_{\mathsf{X}\sim\pi}[\mathsf{X}^4]<\infty.\end{aligned}$$

Thus the dominated convergence theorem yields $\mathbb{E}[(2n)^{-1}\langle\|\sigma-\beta^\star\|^2\rangle]\to\eta_*^{-1}$ almost surely, and Lemma 6 concludes the proof. ■

Proof of Theorem 1.10, Unbounded Support: Having established Theorem 1.9 under Assumption 1.3, the proofs of Proposition 5 and Theorem 1.10 are the same as in the setting where π has bounded support. We note that the expectation of the right side of (53) remains finite and independent of n, by the sub-Gaussian condition (5) for π, justifying the application of the dominated convergence theorem. The same bound (53) holds for $\|r_1^t\|^4/n^2$ and each $t \ge 1$, so that the application of the dominated convergence theorem for (59) is justified by the Lipschitz condition for f and the same argument. ■

Appendix G
Auxiliary Lemmas

A. Empirical Wasserstein Convergence

Definition 7.1: For a matrix $(v_1,\ldots,v_k) = (v_{i,1},\ldots,v_{i,k})_{i=1}^n \in \mathbb{R}^{n\times k}$ and a random vector $(\mathsf{V}_1,\ldots,\mathsf{V}_k)$, we write

$$(v_1,\ldots,v_k)\xrightarrow{W}(\mathsf{V}_1,\ldots,\mathsf{V}_k)$$

for the convergence of the empirical distribution of rows of $(v_1,\ldots,v_k)$ to $(\mathsf{V}_1,\ldots,\mathsf{V}_k)$ in Wasserstein-$\mathfrak{p}$ for every order $\mathfrak{p}\ge1$. This means that $(\mathsf{V}_1,\ldots,\mathsf{V}_k)$ has finite mixed moments of all orders, and for any continuous function $f:\mathbb{R}^k\to\mathbb{R}$ satisfying

$$|f(v_1,\ldots,v_k)|\le C\left(1+\|(v_1,\ldots,v_k)\|^{\mathfrak{p}}\right) \tag{217}$$

for some $C>0$ and $\mathfrak{p}\ge1$, we have $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n f(v_{i,1},\ldots,v_{i,k})=\mathbb{E}[f(\mathsf{V}_1,\ldots,\mathsf{V}_k)]$.

The following results are direct consequences of [50, Propositions E.1, E.2, E.4, F.2].

Proposition 9: Suppose $V\in\mathbb{R}^{n\times t}$ has i.i.d. rows equal in law to $\mathsf{V}\in\mathbb{R}^t$, which has finite mixed moments of all orders. Then $V\xrightarrow{W}\mathsf{V}$ almost surely as $n\to\infty$. Furthermore, if $E\in\mathbb{R}^{n\times k}$ is deterministic with $E\xrightarrow{W}\mathsf{E}$, then $(V,E)\xrightarrow{W}(\mathsf{V},\mathsf{E})$ almost surely where V is independent of E.

Proposition 10: Suppose $V\in\mathbb{R}^{n\times k}$ satisfies $V\xrightarrow{W}\mathsf{V}$ as $n\to\infty$, and $g:\mathbb{R}^k\to\mathbb{R}^l$ is continuous with $\|g(v)\|\le C(1+\|v\|)^{\mathfrak{p}}$ for some $C>0$ and $\mathfrak{p}\ge1$. Then $g(V)\xrightarrow{W}g(\mathsf{V})$ where $g(\cdot)$ is applied row-wise to V.

Proposition 11: Suppose $V\in\mathbb{R}^{n\times k}$, $W\in\mathbb{R}^{n\times l}$, and $M_n, M\in\mathbb{R}^{k\times l}$ satisfy $V\xrightarrow{W}\mathsf{V}$, $W\xrightarrow{W}0$, and $M_n\to M$ entrywise as $n\to\infty$. Then $VM_n+W\xrightarrow{W}\mathsf{V}^\top\cdot M$.

Proposition 12: Fix $l\ge0$, let $O\sim\mathrm{Haar}(\mathbb{SO}(n-l))$, and let $v\in\mathbb{R}^{n-l}$ and $\Pi\in\mathbb{R}^{n\times(n-l)}$ be deterministic, where Π has orthonormal columns and $n^{-1}\|v\|^2\to\sigma^2$ as $n\to\infty$. Then $\Pi Ov\xrightarrow{W}\mathsf{Z}\sim N(0,\sigma^2)$ almost surely. Furthermore, if $E\in\mathbb{R}^{n\times k}$ is deterministic with $E\xrightarrow{W}\mathsf{E}$, then $(\Pi Ov,E)\xrightarrow{W}(\mathsf{Z},\mathsf{E})$ almost surely where Z is independent of E.

We note that Proposition 12 is stated in [50, Proposition F.2(a)] for $O\sim\mathrm{Haar}(\mathbb{O}(n-l))$, but the proof is identical also for $O\sim\mathrm{Haar}(\mathbb{SO}(n-l))$.

B. Properties of Cauchy- and R-Transform

Let $\{\mu_k\}_{k\ge2}$ and $\{\kappa_k\}_{k\ge1}$ be the central moments and free cumulants of $-\mathsf{D}^2$ respectively (see e.g. [51, Lecture 11]). In particular, $\kappa_1=-\mathbb{E}\mathsf{D}^2=-d_*$ and $\kappa_2=\mu_2=\mathbb{V}\left(\mathsf{D}^2\right)$. The following shows that the Cauchy- and R-transforms of $-\mathsf{D}^2$ are well-defined by (10), and reviews their properties.

Lemma 13: Let $G(\cdot)$ and $R(\cdot)$ be the Cauchy- and R-transforms of $-\mathsf{D}^2$ under Assumption 1.1.

(a) The function $G:(-d_-,\infty)\to\mathbb{R}$ is positive and strictly decreasing. Setting $G(-d_-):=\lim_{z\to-d_-}G(z)\in(0,\infty]$, G admits a functional inverse $G^{-1}:(0,G(-d_-))\to(-d_-,\infty)$.

(b) The function $R:(0,G(-d_-))\to\mathbb{R}$ is negative and strictly increasing.

(c) Suppose that Theorem 1.4 holds. Then $G(-d_-)\in[(2\mathfrak{e})^{-1},\infty]$. Furthermore, $\kappa_2\le\min\{\mathfrak{e}^2,d_*\mathfrak{e}\}$, and for $k\ge2$,

$$|\mu_k|\le\mathfrak{e}^{k-2}\kappa_2\le\mathfrak{e}^k,\qquad|\kappa_k|\le16^k\mathfrak{e}^{k-2}\kappa_2\le(16\mathfrak{e})^k. \tag{218}$$

For all $z\in(0,(16\mathfrak{e})^{-1})$, the R-transform admits the convergent series expansion

$$R(z)=\sum_{k\ge1}\kappa_k z^{k-1}. \tag{219}$$

Proof:

(a) The positivity and monotonicity of G follow directly from the definition. Since $\lim_{z\to-d_-}G(z)=G(-d_-)$ and $\lim_{z\to\infty}G(z)=0$, G has an inverse on the stated domain.

(b) For any $z \in (0, G(-d_-))$,

$$z = G(G^{-1}(z)) = \mathbb{E}\frac{1}{\mathsf{D}^2 + R(z) + \frac{1}{z}} < \frac{1}{R(z) + \frac{1}{z}}$$

where the inequality is strict because D^2 has strictly positive mean and variance. Then $R(z) < 0$. Furthermore, by Jensen's inequality $\mathbb{E}\frac{1}{(\mathsf{D}^2+G^{-1}(z))^2} > \left(\mathbb{E}\frac{1}{\mathsf{D}^2+G^{-1}(z)}\right)^2 = z^2$ which also holds strictly since D^2 has strictly positive variance,

$$R'(z) = \frac{1}{G'(G^{-1}(z))} + \frac{1}{z^2} = -\left(\mathbb{E}\frac{1}{(\mathsf{D}^2 + G^{-1}(z))^2}\right)^{-1} + \frac{1}{z^2} > 0.$$

Thus $R(z)$ is strictly increasing.

(c) Under Theorem 1.4, we have both $d_- \in [d_* - \mathfrak{e}, d_* + \mathfrak{e}]$ and $\mathsf{D}^2 \in [d_* - \mathfrak{e}, d_* + \mathfrak{e}]$ almost surely, so $G(-d_-) \in [(2\mathfrak{e})^{-1}, \infty]$. Furthermore $\kappa_2 = \mathbb{V}(\mathsf{D}^2) \le \mathfrak{e}^2$, and also $\kappa_2 \le \mathbb{E}\mathsf{D}^2(d_* + \mathfrak{e}) - (\mathbb{E}\mathsf{D}^2)^2 = d_*\mathfrak{e}$. For any $k \ge 2$,

$$|\mu_k| = |\mathbb{E}[(-\mathsf{D}^2 + d_*)^k]| \le \mathfrak{e}^{k-2}\mathbb{E}\left(-\mathsf{D}^2 + d_*\right)^2 = \mathfrak{e}^{k-2}\kappa_2 \le \mathfrak{e}^k.$$

The free cumulants κ_k for $k \ge 2$ are the same as those of the centered variable $-\mathsf{D}^2 + d_*$. Then, setting $\mu_1 = 0$, the non-crossing moment-cumulant relations applied to $-\mathsf{D}^2 + d_*$ yield

$$|\kappa_k| = \left|\sum_{\pi\in\mathrm{NC}(k)} \mathrm{Mobi}(\pi, 1_k) \cdot \prod_{S\in\pi} \mu_{|S|}\right| \le 16^k \max_{\pi\in\mathrm{NC}(k)} \prod_{S\in\pi} |\mu_{|S|}| \le 16^k \mathfrak{e}^{k-2}\kappa_2 \le (16\mathfrak{e})^k$$

where $\mathrm{NC}(k)$ is the lattice of all non-crossing partitions of $\{1, \dots, k\}$, $\mathrm{Mobi}(\cdot,\cdot)$ are the Möbius functions on the non-crossing partition lattice, 1_k is the trivial partition consisting of the single set $\{1, \dots, k\}$, and the first inequality applies $|\mathrm{Mobi}(\pi, 1_k)| \le 4^k$ and $|\mathrm{NC}(k)| \le 4^k$ [51, Proposition 13.15]. The statement on the R-transform follows from [51, Notation 12.6, Proposition 13.15]). ∎

C. Prior Distribution

We verify that Assumption 1.3 holds for priors having bounded support or log-concave density.

Proposition 13: Suppose π has mean 0 and variance $\rho_* > 0$. Then Assumption 1.3 holds if

(a) π has support contained in $[-\sqrt{\mathfrak{C}}, \sqrt{\mathfrak{C}}]$, or

(b) π admits a Lebesgue density function $e^{-g(x)}$ for all $x \in \mathbb{R}$, where $g''(x) \ge 1/\mathfrak{C}$.

Proof: Under (a), both statements of (6) and the first statement of (5) are evident, and the second statement of (5) follows from Hoeffding's inequality. Under (b), the first statement of (5) follows from the Brascamp-Lieb inequality $\mathbb{V}[f(\mathsf{X}^\star)] \le \mathfrak{C} \cdot \mathbb{E}[f'(\mathsf{X}^\star)^2]$, see e.g. [52, Theorem 13.13], and the second from the Bakry-Emery theorem, see e.g. [52, Theorem 13.6, Proposition 13.8]. We observe that for any $\Gamma \preceq (4\mathfrak{C})^{-1}I$, the measure μ has a density $e^{-g_\mu(x)}$ where $\nabla^2 g_\mu(x) \succeq 3/(4\mathfrak{C})$. Hence the Brascamp-Lieb inequality applies also to μ, and both statements of (6) follow. ∎

In the proofs of the main results, we use the following implications of Assumption 1.3.

Proposition 14: Under Assumption 1.3, the posterior mean denoiser $f(y, \gamma)$ defined by (8) is continuously-differentiable and Lipschitz in y, with derivative $\frac{\partial}{\partial y} f(y, \gamma) = \gamma \cdot \mathbb{V}_\mu[x]$. In particular, $|\frac{\partial}{\partial y} f(y, \gamma)| \le C\gamma$ for a constant $C > 0$ depending only on $\mathfrak{C}$.

Proof: In the notation of Assumption 1.3, setting $k = 1$, $\Gamma = -\frac{\gamma}{2}$, and $z = \gamma y$, we have $f(y, \gamma) = \langle x \rangle_\mu$. A straightforward application of the dominated convergence theorem shows that $y \mapsto f(y, \gamma)$ is continuously differentiable, with derivative $\frac{\partial}{\partial y} f(y, \gamma) = \gamma \cdot \mathbb{V}_\mu[x]$. The inequality $|\frac{\partial}{\partial y} f(y, \gamma)| \le C\gamma$ then follows from (6), and this implies that $f(y, \gamma)$ is $(C\gamma)$-Lipschitz in y. ∎

Proposition 15: Under Assumption 1.3, suppose $\sigma_1, \dots, \sigma_n \overset{iid}{\sim} \pi$. Then for a universal constant $c_0 > 0$ and any $s > 0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n(\sigma_i^2 - \rho_*)\right| \ge s\right) \le 2\exp\left(-c_0 \min\left(\frac{s^2}{\mathfrak{C}^2}, \frac{s}{\mathfrak{C}}\right) n\right)$$

Proof: Under the sub-Gaussian condition (5), the random variables σ_i^2 are sub-exponential with mean ρ_*, so the result follows from Bernstein's inequality, see e.g. [53, Theorem 2.8.2]. ∎

Proposition 16: Fix $k \in \{1, 2\}$ and let Γ, z and $\langle\cdot\rangle_\mu, \mathbb{V}_\mu[\cdot]$ be as defined in Assumption 1.3. Let $\gamma_{\max}, \gamma_{\min}$ be the largest and smallest eigenvalues of Γ. Then for any unit vector $v \in \mathbb{R}^k$ and for a constant $C > 0$ depending only on $\mathfrak{C}$,

$$\mathbb{V}_\mu[(v^\top x)^2] \le C\left(1 + \frac{\|z\|^2}{(\mathfrak{C}^{-1} - \gamma_{\max})^2} + \frac{\max(-\gamma_{\min}, 0)}{\mathfrak{C}^{-1} - \gamma_{\max}}\right).$$

Proof: Applying both conditions of (6),

$$\mathbb{V}_\mu[(v^\top x)^2] \le C(1 + \mathbb{V}_\mu[v^\top x] + \langle v^\top x\rangle_\mu^2) \le C(1 + C + \langle v^\top x\rangle_\mu^2) \tag{220}$$

so it suffices to bound $\langle v^\top x\rangle_\mu^2$. We apply an idea similar to [54, Proposition 2]: Set $a = 3/(4\mathfrak{C})$ and denote $\Omega = aI - 2\Gamma$, $w = \Omega^{-1}z$, $\|x - w\|_\Omega^2 = (x - w)^\top\Omega(x - w)$, and $d\pi(x) = \prod_{i=1}^k d\pi(x_i)$. Let $\omega_{\min} = a - 2\gamma_{\max}$ be the smallest eigenvalue of Ω, and note that $\omega_{\min} > 0$ because $\gamma_{\max} < (4\mathfrak{C})^{-1}$ in Assumption 1.3. We have

$$\langle v^\top x\rangle_\mu = \frac{\int v^\top x \cdot e^{\frac{a}{2}\|x\|^2} e^{-\frac{1}{2}\|x-w\|_\Omega^2} d\pi(x)}{\int e^{\frac{a}{2}\|x\|^2} e^{-\frac{1}{2}\|x-w\|_\Omega^2} d\pi(x)}.$$

Denote $c_\Omega(w) = \log \int e^{\frac{a}{2}\|x\|^2} e^{-\frac{1}{2}\|x-w\|_\Omega^2} d\pi(x)$. On the event $\|x - w\|_\Omega^2 \ge -2c_\Omega(w)$, we have

$$\frac{e^{-\frac{1}{2}\|x-w\|_\Omega^2}}{\int e^{\frac{a}{2}\|x\|^2} e^{-\frac{1}{2}\|x-w\|_\Omega^2} d\pi(x)} \le 1.$$

On the complementary event $\|x-w\|_\Omega^2 < -2c_\Omega(w)$, we have $c_\Omega(w) < 0$ and

$$\begin{aligned}|v^\top x| &\le \|w\| + \|x-w\| \\ &\le \|w\| + \omega_{\min}^{-1/2}\|x-w\|_\Omega \\ &\le \omega_{\min}^{-1}\|z\| + \omega_{\min}^{-1/2}(-2c_\Omega(w))^{1/2}.\end{aligned}$$

Thus, combining these bounds,

$$\begin{aligned}|\langle v^\top x\rangle_\mu| \le &\int |v^\top x| \cdot e^{\frac{a}{2}\|x\|^2} \cdot 1\, d\pi(x) \\ &+ \Big(\omega_{\min}^{-1}\|z\| + \omega_{\min}^{-1/2}(-2c_\Omega(w))^{1/2}\Big) \cdot \mathbf{1}\{c_\Omega(w) < 0\}.\end{aligned}$$

Applying $\frac{a}{2} = \frac{3}{8\mathfrak{C}}$ and the sub-Gaussian tail bound (5), it is easily checked that the first term satisfies $\int |v^\top x|\, e^{\frac{a}{2}\|x\|^2} d\pi(x) < C$ for a constant $C > 0$ depending only on $\mathfrak{C}$. For the second term, by Jensen's inequality and the condition that π has mean 0,

$$\begin{aligned}-c_\Omega(w) &\le \int \left(-\frac{a}{2}\|x\|^2 + \frac{1}{2}\|x-w\|_\Omega^2\right) d\pi(x) \\ &\le -\gamma_{\min}\int \|x\|^2\, d\pi(x) + \frac{1}{2}\|w\|_\Omega^2 \\ &\le C\Big(\max(-\gamma_{\min}, 0) + \omega_{\min}^{-1}\|z\|^2\Big).\end{aligned}$$

Applying this above yields $|\langle v^\top x\rangle_\mu| \le C(1 + \omega_{\min}^{-1}\|z\| + \omega_{\min}^{-1/2}\max(-\gamma_{\min}, 0)^{1/2})$. Then, applying this to (220) and using $\omega_{\min} \ge c(\mathfrak{C}^{-1} - \gamma_{\max})$ for a universal constant $c > 0$ concludes the proof. ■

D. Varadhan's Lemma

We apply the following version of Varadhan's lemma; the proof is a straightforward extension of [47, Lemmas 4.3.4 and 4.3.6] and omitted for brevity.

Lemma 14: Let $(X_n)_{n\ge1}$ be a sequence of random variables taking values in a regular topological space $\mathcal{X}$, and let $f : \mathcal{X} \to \mathbb{R}$ be a bounded continuous function.

(a) If $\lambda^* : \mathcal{X} \to [0, \infty]$ is such that $\liminf_{n\to\infty} n^{-1}\log \mathbb{P}[X_n \in G] \ge -\inf_{x\in G}\lambda^*(x)$ for all open $G \subseteq \mathcal{X}$, then

$$\begin{aligned}&\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{E}[\mathbb{I}(X_n \in G)e^{nf(X_n)}] \\ &\quad\ge \sup_{x\in G} f(x) - \lambda^*(x) \text{ for all open } G \subseteq \mathcal{X}.\end{aligned}$$

(b) If $\lambda^* : \mathcal{X} \to [0, \infty]$ is lower-semicontinuous, the level sets $\{x \in \mathcal{X} : \lambda^*(x) \le K\}$ are compact for all $K \in [0, \infty)$, and $\limsup_{n\to\infty} n^{-1}\log \mathbb{P}[X_n \in F] \le -\inf_{x\in F}\lambda^*(x)$ for all closed $F \subseteq \mathcal{X}$, then

$$\begin{aligned}&\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{E}[\mathbb{I}(X_n \in F)e^{nf(X_n)}] \\ &\quad\le \sup_{x\in F} f(x) - \lambda^*(x) \text{ for all closed } F \subseteq \mathcal{X}.\end{aligned}$$

E. Extension of Results to $O \sim \mathrm{Haar}(\mathbb{O}(n))$

We explain the claim of Remark 1.6. Observe first that if D, Q are random and independent of $O, \beta^\star, \epsilon$, where (4) holds almost surely as $n, m \to \infty$, then the results of Theorems 1.7, 1.9, and 1.10 all hold almost surely as $n, m \to \infty$ conditional on D, Q, and hence also unconditionally. If $A = Q^\top DO$ where $O \sim \mathrm{Haar}(\mathbb{O}(n))$, then we have the equality in law $O \stackrel{L}{=} PO'$ where $O' \sim \mathrm{Haar}(\mathbb{SO}(n))$ and $P = \mathrm{diag}(1, \ldots, 1, b) \in \mathbb{R}^{n\times n}$ with $b \in \{+1, -1\}$ having equal probability. If $n > m$, then $DP = D$, so $A = Q^\top DO'$ and the model is identical to the setting of $O \sim \mathrm{Haar}(\mathbb{SO}(n))$. If $n \le m$, then $DP = P'D$ where $P' = \mathrm{diag}(1, \ldots, 1, b, 1, \ldots, 1) \in \mathbb{R}^{m\times m}$ has b in the n^{th} entry. Setting $Q' = P'Q$, this implies $A = (Q')^\top DO'$. The asymptotic statements of Theorems 1.7, 1.9, and 1.10 thus hold almost surely conditional on b, and hence also unconditionally.

Acknowledgment

The authors would like to thank the Associate Editor and the reviewers for their helpful comments.

References

[1] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.

[2] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269–280, Jan. 2010.

[3] Y. Guan and M. Stephens, "Bayesian variable selection regression for genome-wide association studies and other large-scale problems," *Ann. Appl. Statist.*, vol. 5, no. 3, pp. 1780–1815, Sep. 2011.

[4] D. Guo and S. Verdu, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, Jun. 2005.

[5] A. Montanari and D. Tse, "Analysis of belief propagation for non-linear problems: The example of CDMA (or: How to prove Tanaka's formula)," in *Proc. IEEE Inf. Theory Workshop*, Mar. 2006, pp. 160–164.

[6] J. Barbier, M. Dia, N. Macris, and F. Krzakala, "The mutual information in random linear estimation," in *Proc. 54th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2016, pp. 625–632.

[7] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for compressed sensing with Gaussian matrices is exact," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 665–669.

[8] J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová, "Optimal errors and phase transitions in high-dimensional generalized linear models," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 12, pp. 5451–5460, Mar. 2019.

[9] J. Barbier and N. Macris, "The adaptive interpolation method: A simple scheme to prove replica formulas in Bayesian inference," *Probab. Theory Rel. Fields*, vol. 174, nos. 3–4, pp. 1133–1185, Aug. 2019.

[10] J. Barbier, N. Macris, M. Dia, and F. Krzakala, "Mutual information and optimality of approximate message-passing in random linear estimation," *IEEE Trans. Inf. Theory*, vol. 66, no. 7, pp. 4270–4303, Jul. 2020.

[11] J. Qiu and S. Sen, "The TAP free energy for high-dimensional linear regression," *Ann. Appl. Probab.*, vol. 33, no. 4, pp. 2643–2680, Aug. 2023.

[12] K. Takeda, S. Uda, and Y. Kabashima, "Analysis of CDMA systems that are characterized by eigenvalue spectrum," *Europhys. Lett. (EPL)*, vol. 76, no. 6, pp. 1193–1199, Dec. 2006.

[13] A. M. Tulino, G. Caire, S. Verdu, and S. Shamai (Shitz), "Support recovery with sparsely sampled free random matrices," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4243–4271, Jul. 2013.

[14] J. Barbier, N. Macris, A. Maillard, and F. Krzakala, "The mutual information in random linear estimation beyond i.i.d matrices," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 1390–1394.

[15] M. Opper and O. Winther, "Adaptive and self-averaging thouless-anderson-palmer mean-field theory for probabilistic modeling," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 5, Oct. 2001, Art. no. 056131.

[16] M. Opper and O. Winther, "Tractable approximations for probabilistic models: The adaptive thouless-anderson-palmer mean field approach," *Phys. Rev. Lett.*, vol. 86, no. 17, pp. 3695–3699, Apr. 2001.

[17] M. Opper, O. Winther, and M. J. Jordan, "Expectation consistent approximate inference," *J. Mach. Learn. Res.*, vol. 6, no. 12, pp. 1–28, 2005.

[18] P. Schniter, S. Rangan, and A. K. Fletcher, "Vector approximate message passing for the generalized linear model," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2016, pp. 1525–1529.
[19] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, 2017.
[20] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 501–505.
[21] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6664–6684, Oct. 2019.
[22] K. Takeuchi, "Bayes-optimal convolutional AMP," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 1385–1390.
[23] L. Liu, S. Huang, and B. M. Kurkoski, "Memory AMP," *IEEE Trans. Inf. Theory*, vol. 68, no. 12, pp. 8015–8039, Dec. 2022.
[24] A. Maillard, L. Foini, A. L. Castellanos, F. Krzakala, M. Mézard, and L. Zdeborová, "High-temperature expansions and message passing algorithms," *J. Stat. Mech., Theory Exp.*, vol. 2019, no. 11, Nov. 2019, Art. no. 113301.
[25] T. Takahashi and Y. Kabashima, "Macroscopic analysis of vector approximate message passing in a model-mismatched setting," *IEEE Trans. Inf. Theory*, vol. 68, no. 8, pp. 5579–5600, Aug. 2022.
[26] E. Marinari, G. Parisi, and F. Ritort, "Replica field theory for deterministic models. II. A non-random spin glass with glassy behaviour," *J. Phys. A, Math. Gen.*, vol. 27, no. 23, pp. 7647–7668, Dec. 1994.
[27] G. Parisi and M. Potters, "Mean-field equations for spin models with orthogonal interaction matrices," *J. Phys. A, Math. Gen.*, vol. 28, no. 18, pp. 5267–5285, Sep. 1995.
[28] C. Gerbelot, A. Abbara, and F. Krzakala, "Asymptotic errors for high-dimensional convex penalized linear regression beyond Gaussian matrices," in *Proc. Conf. Learn. Theory*, 2020, pp. 1682–1713.
[29] C. Gerbelot, A. Abbara, and F. Krzakala, "Asymptotic errors for teacher–student convex generalized linear models (or: How to prove Kabashima's replica formula)," *IEEE Trans. Inf. Theory*, vol. 69, no. 3, pp. 1824–1852, Mar. 2023.
[30] R. Dudeja, Y. M. Lu, and S. Sen, "Universality of approximate message passing with semirandom matrices," *Ann. Probab.*, vol. 51, no. 5, pp. 1616–1683, Sep. 2023.
[31] T. Wang, X. Zhong, and Z. Fan, "Universality of approximate message passing algorithms and tensor networks," 2022, *arXiv:2206.13037*.
[32] R. Dudeja, S. Sen, and Y. M. Lu, "Spectral universality of regularized linear regression with nearly deterministic sensing matrices," 2022, *arXiv:2208.02753*.
[33] Z. Fan and Y. Wu, "The replica-symmetric free energy for Ising spin glasses with orthogonally invariant couplings," 2021, *arXiv:2105.02797*.
[34] Z. Fan, Y. Li, and S. Sen, "TAP equations for orthogonally invariant spin glasses at high temperature," 2022, *arXiv:2202.09325*.
[35] E. Bolthausen, "A Morita type proof of the replica-symmetric formula for SK," in *Proc. Int. Conf. Stat. Mech. Classical Disordered Syst.* Cham, Switzerland: Springer, 2018, pp. 63–93.
[36] J. Ding and N. Sun, "Capacity lower bound for the Ising perceptron," in *Proc. 51st Annu. ACM SIGACT Symp. Theory Comput.*, 2019, pp. 816–827.
[37] E. Bolthausen, S. Nakajima, N. Sun, and C. Xu, "Gardner formula for Ising perceptron models at small densities," in *Proc. Conf. Learn. Theory*, 2022, pp. 1787–1911.
[38] D. Guo, S. Shamai (Shitz), and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.
[39] J. Barbier, "Overlap matrix concentration in optimal Bayesian inference," *Inf. Inference, A J. IMA*, vol. 10, no. 1, pp. 597–623, Oct. 2020.
[40] J. Barbier and D. Panchenko, "Strong replica symmetry in high-dimensional optimal Bayesian inference," *Commun. Math. Phys.*, vol. 393, no. 3, pp. 1199–1239, Aug. 2022.
[41] F. Guerra, "Broken replica symmetry bounds in the mean field spin glass model," *Commun. Math. Phys.*, vol. 233, no. 1, pp. 1–12, Feb. 2003.
[42] S. B. Korada and N. Macris, "Tight bounds on the capacity of binary input random CDMA systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5590–5613, Nov. 2010.
[43] G. Reeves, "Additivity of information in multilayer networks via additive Gaussian noise transforms," in *Proc. 55th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2017, pp. 1064–1070.
[44] D. J. Thouless, P. W. Anderson, and R. G. Palmer, "Solution of 'solvable model of a spin glass,'" *Phil. Mag.*, vol. 35, no. 3, pp. 593–601, Mar. 1977.
[45] K. Takeuchi, "On the convergence of orthogonal/vector AMP: Long-memory message-passing strategy," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2022, pp. 1366–1371.
[46] M. Payaro and D. P. Palomar, "Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3613–3628, Aug. 2009.
[47] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, vol. 38. Berlin, Germany: Springer, 1998.
[48] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 2015.
[49] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*, vol. 118. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[50] Z. Fan, "Approximate message passing algorithms for rotationally invariant matrices," *Ann. Statist.*, vol. 50, no. 1, pp. 197–224, Feb. 2022.
[51] A. Nica and R. Speicher, *Lectures on the Combinatorics of Free Probability*, vol. 13. Cambridge, U.K.: Cambridge Univ. Press, 2006.
[52] L. Erdős and H.-T. Yau, *A Dynamical Approach to Random Matrix Theory*, vol. 28. Providence, RI, USA: American Mathematical Society, 2017.
[53] R. Vershynin, *High-Dimensional Probability: An Introduction With Applications in Data Science*, vol. 47. Cambridge, U.K.: Cambridge Univ. Press, 2018.
[54] Y. Polyanskiy and Y. Wu, "Wasserstein continuity of entropy and outer bounds for interference channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3992–4002, Jul. 2016.

Yufan Li received the B.A.Sc. degree from the University of Toronto in 2018 and the M.E. degree from Harvard University in 2020, where he is currently pursuing the Ph.D. degree with the Department of Statistics. His research interests include high dimensional statistics, statistical physics, and machine learning.

Zhou Fan received the A.B./S.M. degree from Harvard University in 2010, the M.A.St. degree from the University of Cambridge in 2011, and the Ph.D. degree from Stanford University in 2018. He is currently an Assistant Professor with the Department of Statistics and Data Science, Yale University. His research interests include high-dimensional statistics, random matrix theory and statistical physics, and applications in the biological sciences. He was a recipient of the NSF CAREER Award in 2022.

Subhabrata Sen received the B.Stat. and M.Stat. degrees from the Indian Statistical Institute in 2011 and 2013, respectively, and the Ph.D. degree from Stanford University in 2017. He is currently an Assistant Professor in statistics with Harvard University. He was a recipient of the NSF CAREER Award in 2023. His research interests include high-dimensional statistics, statistical physics, and network analysis.

Yihong Wu received the B.E. degree from Tsinghua University in 2006 and the Ph.D. degree from Princeton University in 2011. He is currently a Professor with the Department of Statistics and Data Science, Yale University. His research interests include the theoretical and algorithmic aspects of high-dimensional statistics, information theory, and optimization. He was a recipient of the NSF CAREER Award in 2017 and the Sloan Research Fellowship in Mathematics in 2018.