Spectral estimators for structured generalized linear models via approximate message passing

Yihan Zhang, Hong Chang Ji, Ramji Venkataramanan, and Marco Mondelli

Abstract. We consider the problem of parameter estimation in a high-dimensional generalized linear model. Spectral methods obtained via the principal eigenvector of a suitable data-dependent matrix provide a simple yet surprisingly effective solution. However, despite their wide use, a rigorous performance characterization, as well as a principled way to preprocess the data, are available only for unstructured (i.i.d. Gaussian and Haar orthogonal) designs. In contrast, real-world data matrices are highly structured and exhibit non-trivial correlations. To address the problem, we consider correlated Gaussian designs capturing the anisotropic nature of the features via a covariance matrix Σ. Our main result is a precise asymptotic characterization of the performance of spectral estimators. This allows us to identify the optimal preprocessing that minimizes the number of samples needed for parameter estimation. Surprisingly, such preprocessing is universal across a broad set of designs, which partly addresses a conjecture on optimal spectral estimators for rotationally invariant models. Our principled approach vastly improves upon previous heuristic methods, including for designs common in computational imaging and genetics. The proposed methodology, based on approximate message passing, is broadly applicable and opens the way to the precise characterization of spiked matrices and of the corresponding spectral methods in a variety of settings.

Contents

1. Introduction

This paper considers the prototypical problem of learning a parameter vector from observations obtained via a generalized linear model (GLM) [61]:

$$y_i = q\big(\langle x_i, \beta^* \rangle, \varepsilon_i\big), \quad 1 \le i \le n, \tag{1.1}$$

where $\beta^* \in \mathbb{R}^d$ consists of (unknown) regression coefficients. The statistician wishes to estimate β^* based on the observations $y = (y_i)_{i=1}^n \in \mathbb{R}^n$ and the covariate vectors $x_1, \ldots, x_n \in \mathbb{R}^d$. The vector $\varepsilon = (\varepsilon_i)_{i=1}^n \in \mathbb{R}^n$ contains (unknown) i.i.d. random variables accounting for noise in the measurements. The (known) link function $q: \mathbb{R}^2 \to \mathbb{R}$ is applied element-wise, i.e., $q(g, \varepsilon) = (q(g_1, \varepsilon_1), \ldots, q(g_n, \varepsilon_n))$ for any $g, \varepsilon \in \mathbb{R}^n$. The non-linearity q generalizes linear regression ($q(g, \varepsilon) = g + \varepsilon$) and incorporates various problems in statistics, machine learning, signal processing and computational biology, e.g., phase retrieval ($q(g, \varepsilon) = |g| + \varepsilon$) (see [39]), 1-bit compressed sensing ($q(g, \varepsilon) = \text{sign}(g) + \varepsilon$) (see [9]), and logistic regression [83].

For estimation in GLMs, several works have considered methods based on convex programming, e.g., [15, 84, 88]. However, these methods often become computationally infeasible as d grows. Thus, fast iterative methods including alternating minimization [69], approximate message passing [76], Wirtinger flow [14], iterative projections [50], and the Kaczmarz method [91] has been developed. Due to their iterative nature, to converge to an informative solution, these procedures require a "warm start", i.e., a vector $\widehat{\beta} \in \mathbb{R}^d$ whose "overlap" $|\langle \widehat{\beta}, \beta^* \rangle| / (\|\widehat{\beta}\|_2 \|\beta^*\|_2)$ with β^* is non-vanishing for large d. In this paper, we focus on *spectral estimators* [22], which provide a simple yet effective approach for estimating β^*, and serve as a warm start for the local methods above. Spectral estimators have been applied in a range of problems including polynomial learning [20], estimation from mixed linear regression [95] and ranking [23]. For the GLM in (1.1), the spectral estimator processes the observations via a function $\mathcal{T}: \mathbb{R} \to \mathbb{R}$ and outputs the principal eigenvector of the matrix

$$D := \sum_{i=1}^n x_i x_i^\top \mathcal{T}(y_i) \in \mathbb{R}^{d \times d}. \tag{1.2}$$

To understand the accuracy of spectral estimators, it is crucial to: (i) characterize their performance (e.g., in terms of limiting overlap), and (ii) design the preprocessing

function $\mathcal{T}$ that minimizes the sample complexity, i.e., the number n of observations required to attain a desired limiting overlap. This work gives precise answers to both these questions, providing solid performance guarantees as well as a principled basis for optimizing spectral estimators used in practical applications.

A line of work [14, 21, 69] has bounded the sample complexity of spectral estimators obtained from (1.2) for i.i.d. Gaussian designs via matrix concentration inequalities. However, these bounds require the number n of observations to substantially exceed the parameter dimension d, and they are not sharp enough to optimize $\mathcal{T}$. Using tools from random matrix theory, the works [55, 62] obtained tight results in the proportional regime where $n, d \to \infty$ and $n/d \to \delta$ for a fixed constant $\delta \in (0, \infty)$ (called the "aspect ratio"). Specifically, a *phase transition* phenomenon is established: if δ exceeds a critical value (referred to as the "spectral threshold"), then (i) a spectral gap emerges between the first two eigenvalues of D, and (ii) the spectral estimator attains non-vanishing correlation with β^*. For δ below this critical value, there is no outlier to the right of the spectrum of D, and the spectral estimator is asymptotically independent of β^*. This precise characterization allows to derive the optimal preprocessing function that minimizes the spectral threshold [62] and also that maximizes the overlap for a given δ (see [56]). These results are extended by [33, 57] to cover a sub-sampled Haar design, consisting of a subset of columns from a uniformly random orthogonal matrix.

The line of work above crucially relies on the design matrix $X = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top$ being unstructured, namely i.i.d. Gaussian or rotationally-invariant with unit singular values. In contrast, design matrices occurring in practice are highly structured and their entries exhibit significant correlations (e.g., in computational genomics [54] and imaging [13]). In this paper, we capture the correlation and heterogeneity of the data via general (*correlated*) Gaussian designs. Specifically, each covariate x_i is an i.i.d. d-dimensional zero-mean Gaussian vector with an arbitrary positive definite covariance matrix $\Sigma/n \in \mathbb{R}^{d \times d}$. The covariance matrix Σ captures correlations between covariates and the heterogeneity in their variances. General Gaussian designs (e.g., with Toeplitz or circulant covariance structures) have been widely adopted in high-dimensional regression models [43–45, 85, 87, 97]. However, existing results largely focus on (penalized) maximum-likelihood estimators for linear and logistic models [16, 17, 82, 83, 100]. An asymptotic theory of spectral estimators for GLMs with general Gaussian designs has been lacking. One significant challenge is that current techniques for i.i.d. and Haar designs all crucially depend on their right rotational invariance, which fails to hold for correlated covariates.

1.1. Main results

Our main contribution is to give a *precise asymptotic characterization of the overlap between the leading eigenvector of D and the unknown parameter β^**, as well as the locations of the top two eigenvalues of D, provided a criticality condition holds. This is the content of Theorem 3.1, which is informally stated below.

Theorem (Informal version of Theorem 3.1). *Consider the GLM given in (1.1) under a general Gaussian design with covariance Σ/n. Let us assume that $n, d \to \infty$ with $n/d \to \delta \in (0, \infty)$. Let $\bar{\Sigma}$ be a random variable whose law is the limiting eigenvalue distribution of Σ. Fix $\mathcal{T}: \mathbb{R} \to \mathbb{R}$ and let β^{spec} denote the leading eigenvector of the matrix D defined in (1.2). Then, there exist computable scalars $F(\delta, \bar{\Sigma}, \mathcal{T})$, $\lambda_1(\delta, \bar{\Sigma}, \mathcal{T})$, $\lambda_2(\delta, \bar{\Sigma}, \mathcal{T})$, $\eta(\delta, \bar{\Sigma}, \mathcal{T})$ such that the following holds. If $F(\delta, \bar{\Sigma}, \mathcal{T}) > 0$, then:*

(1) *the limits of the top two eigenvalues of D equal $\lambda_1(\delta, \bar{\Sigma}, \mathcal{T}) > \lambda_2(\delta, \bar{\Sigma}, \mathcal{T})$, respectively; and*

(2) $|\langle \beta^{\mathrm{spec}}, \beta^* \rangle| / (\|\beta^{\mathrm{spec}}\|_2 \|\beta^*\|_2) \to \eta(\delta, \bar{\Sigma}, \mathcal{T}) > 0.$

The performance characterization of spectral estimators provided by our main result opens the way to their principled optimization. In Section 3.1, we optimize the preprocessing $\mathcal{T}$ towards minimizing the spectral threshold. A remarkable feature of the optimal preprocessing is that it depends on the covariance matrix Σ of the design only through its normalized trace. In other words, it is *universally optimal over any covariance structure* with fixed trace. An important practical implication is that to apply the optimal spectral estimator, only the normalized trace $\frac{1}{d} \operatorname{Tr}(\Sigma)$ needs to be estimated, instead of the whole matrix Σ. In the proportional regime, the scalar $\frac{1}{d} \operatorname{Tr}(\Sigma)$ can be estimated consistently using a simple plug-in estimator involving the sample covariance matrix. In contrast, consistent estimation of Σ typically requires a sample size larger than that needed by the spectral estimator itself, see Remark 3.7 for details. Our result on the optimal spectral threshold also resolves in part a conjecture in [59] on optimal spectral methods for rotationally invariant designs; see Section 3.2.

The criticality condition $F(\delta, \bar{\Sigma}, \mathcal{T}) > 0$ does not depend on the data and can be easily checked numerically. Whenever the condition holds, our results imply that (i) the top eigenvalue is detached from the bulk of the spectrum of D, hence constituting an outlier, and (ii) the spectral estimator attains strictly positive asymptotic overlap. We conjecture that $F(\delta, \bar{\Sigma}, \mathcal{T}) > 0$ is in fact necessary to achieve positive overlap, see Remark 3.4.

1.2. Technical ideas

Our goal is to characterize top eigenvector and top two eigenvalues of the matrix D in (1.2), which can be expressed as $X^\top T X$, with $X = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top \in \mathbb{R}^{n \times d}$ and $T = \text{diag}(\mathcal{T}(y)) \in \mathbb{R}^{n \times n}$. From the analysis for i.i.d. Gaussian designs [55, 62], we expect that the dependence between T and X will, under a suitable criticality condition, lead to an outlier eigenvalue in the spectrum of D, and when this happens, the corresponding eigenvector (i.e., the spectral estimator) has non-zero overlap with β^*. Note that

$$D = X^\top T X = \Sigma^{1/2} \widetilde{X}^\top T \widetilde{X} \Sigma^{1/2},$$

where $\widetilde{X} \in \mathbb{R}^{n \times d}$ has i.i.d. $\mathcal{N}(0, 1/n)$ entries. If T were independent of X, then D would be a *spiked separable covariance matrix* recently studied in [28]. However, in the GLM setting, y (and, thus, T) depends on X via the 1-dimensional projection $X\beta^*$, so results from [28] cannot be applied. Indeed, to the best of our knowledge, there is no off-the-shelf result in random matrix theory giving spectral information on D. Existing techniques for i.i.d. Gaussian designs [55, 62] also seem difficult to extrapolate as X is not isotropic.

To overcome these difficulties, we propose a novel proof strategy using the theory of approximate message passing (AMP). Specifically, AMP refers to a family of iterative algorithms that are specified by a sequence of "denoising" functions. A key feature of AMP is the presence of a memory term, which debiases the iterates, ensuring that their joint empirical distribution is asymptotically Gaussian. This in turn allows to track their covariance structure via a low-dimensional recursion known as *state evolution* [5, 8]. Our key idea is to simulate a power iteration using AMP: via a judicious choice of denoisers, we ensure that the AMP recursion, once executed for a sufficiently large number of steps, approximates an eigenequation of D. Then, we leverage state evolution to:

- identify the location of the outliers in the spectrum of D, by controlling the ℓ_2-norm of the iterates of AMP, and

- establish the limiting correlation between the top eigenvector of D and β^*, by characterizing the inner product of the iterates with the parameter vector β^*.

The idea of using AMP to simulate an algorithm whose output is aligned with the estimator of interest has been used to characterize the asymptotic performance in many settings [10, 11, 32, 52, 78, 83]. We highlight that, for the study of spectral estimators for GLMs, all previous works using AMP as a proof technique [63, 66, 99] require precise knowledge of when a spectral gap emerges. For the settings considered in those works, complete characterizations of the spectrum are available via known results from random matrix theory. This is however not the case for our setting with a correlated Gaussian design. In this work, we exploit random matrix theory tools for

studying the right edge of the bulk. The fundamental novelty of our approach is that the more challenging task of locating the spike is accomplished by AMP.

1.3. Related work

Spectral methods. Spectral methods find applications in various domains across statistics and data science [22] and, as discussed earlier, the spectrally-initialized optimization paradigm is widely employed for estimation from GLMs and their variants. Beyond GLMs, other applications include community detection [1], clustering [70], angular synchronization in cryo-EM [80], inference of low-rank matrices [65] and tensors [64].

Approximate message passing. Approximate message passing algorithms were first proposed for linear regression [31, 46, 48], and have since been applied to several statistical estimation problems, including parameter recovery in a GLM [4, 76, 77, 83, 86]; see the review [40] and references therein. In this paper, AMP is used solely as a tool for analyzing spectral estimators. Following [18, 66, 68], we expect that our results can be used to analyze general first order iterative methods (including AMP itself) with spectral initialization. An alternative way to initialize first order methods is via random initialization. A recent line of work [49, 51] analyzes AMP with spectral and random initializations in the context of symmetric rank-1 matrix estimation, by establishing a non-asymptotic state evolution result. A different non-asymptotic analysis of AMP, leveraging a leave-one-out approach, was recently put forward in [3].

Random matrix theory. The separable covariance matrix model [24, 73, 94] and its spiked counterpart [28, 29] are related to the matrix D that we study, but as discussed earlier, the results in these papers cannot be applied to GLMs with correlated designs. A related (and more general) model is considered in [53], where potential outlier eigenvalues/eigenvectors are identified via a deterministic equivalent of the resolvent. However, [53] provides no explicit condition under which these outliers indeed emerge. In comparison, our result locates both the right edge of the spectral bulk and the outlier eigenvalue, yielding an almost sure characterization. Our approach has the advantage of rendering itself ready for initializing iterative procedures.

Label transformation and generative exponent. The preprocessing function in our work corresponds to the label transformation technique used in [20, 25]. In fact, the thresholding filter in [20] is the same as the subset scheme proposed in [89], and it is a special case of the spectral estimators considered in our paper with preprocessing function $\mathcal{T}^{\text{subset}}(y) = \mathbb{1}\{|y| \geq K_{\text{subset}}\}$. Damian et al. [25] extend the analysis to tensor estimators that provide weak recovery guarantees for n super-linear in d (i.e., not in the proportional regime considered in this work). This handles settings where the

optimal spectral threshold identified in Theorem 3.2 (or [62, Theorem 2]) is infinity, which points to the need of having n that grows faster than d. The insight of [25] is that applying the optimal preprocessing function $\mathcal{T}^*$ lowers the information exponent of q to its generative exponent, which equals the information exponent of the functional composition of q with $\mathcal{T}^*$.

2. Preliminaries

2.1. Generalized linear models with general Gaussian designs

Recall that the goal is to estimate the parameter vector $\beta^* \in \mathbb{R}^d$ from observations obtained via the model in (1.1). We write $y = q(X\beta^*, \varepsilon) \in \mathbb{R}^n$ for the observation vector, with the link function q acting component-wise on its inputs. We make the following assumptions on the model:

(A1) $\beta^* \sim P^{\otimes d}$,[1] where P is a distribution on $\mathbb{R}$ with mean 0 and variance 1.

(A2) For $1 \le i \le n$, $x_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0_d, \Sigma/n)$ independent of β^*, where $\Sigma \in \mathbb{R}^{d \times d}$ is deterministic and strictly positive definite with empirical spectral distribution[2] converging weakly to the law of a random variable $\overline{\Sigma}$ compactly supported on $(0, \infty)$. The spectral norm $\|\Sigma\|_2$ is uniformly bounded over d and, for all $\varsigma > 0$, there exists $d_0 \in \mathbb{N}$ such that for all $d \ge d_0$,

$$\operatorname{supp}(\mu_\Sigma) \subset \operatorname{supp}(\overline{\mu}_\Sigma) + [-\varsigma, \varsigma], \tag{2.1}$$

where μ_Σ and $\overline{\mu}_\Sigma$ denote respectively the empirical and limiting spectral distributions of Σ, supp denotes their support and "+" denotes the Minkowski sum.

(A3) $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n) \in \mathbb{R}^n$ is independent of (β^*, X) and has empirical distribution converging in probability in Wasserstein-2 distance to P_ε which is a distribution on $\mathbb{R}$ with bounded second moment.

(A4) We work in the proportional regime where $n, d \to \infty$ with $n/d \to \delta$ for some $\delta \in (0, \infty)$.

Assumption (A1) specifies an i.i.d. prior distribution on the unknown parameter β^*. We remark that our analysis carries over to $\beta^* \sim \operatorname{Unif}(\sqrt{d}\,\mathbb{S}^{d-1})$ (where $\mathbb{S}^{d-1}$ denotes the unit sphere in dimension d), giving the same results as for $P = \mathcal{N}(0, 1)$.

---

[1]For a tuple of distributions $P_1, \ldots, P_k$, $P_1 \otimes \cdots \otimes P_k$ denotes the product distribution with P_i being its i-th marginal. If all P_i's are equal to P, we use the notation $P^{\otimes k}$.

[2]The empirical spectral distribution of a $p \times p$ matrix is a probability measure that assigns weight $1/p$ to a Dirac mass supported at each of the eigenvalues.

Spectral estimators are unable to exploit any prior structure in the parameter vector, since the eigenvectors of the spectral matrix are not a priori guaranteed to obey structures (e.g., binary, sparse or conic) that may be enjoyed by the parameter. In fact, our results are universal with respect to P. We leave it for future work to perform parameter estimation with prior information taken into account. Furthermore, our results can be extended to the setting where β^* has non-i.i.d. prior and in particular can align with eigenvectors of Σ. See Remark 3.5 for the required modifications for such adaptation.

The *general* Gaussian design in assumption (A2) constitutes the major challenge of this work. We highlight that no distributional assumption is imposed on the matrix Σ: this in particular means that X is only *left* rotationally invariant in law. As such, the model falls out of the bi-rotationally invariant ensemble which has recently attracted a flurry of research [12, 37, 59, 86, 90]. The requirement of strict positive definiteness of Σ could be relaxed to positive semidefiniteness with the modification in the proof that Σ^{-1} is replaced with the pseudo-inverse Σ^+ and $\bar{\Sigma}$ is replaced with a proper mixture of δ_0 (where δ_λ is the Dirac delta measure at $\lambda \in \mathbb{R}$) and a certain absolutely continuous (with respect to the Lebesgue measure) probability measure. The assumption on uniform boundedness of $\|\Sigma\|_2$ is technical and is satisfied by many natural covariance structures used in practice, such as Toeplitz or circulant. The condition (2.1) excludes outlier eigenvalues from the spectrum of Σ. Otherwise, it is known that spikes in Σ will result in spikes in D (see [6, 27, 28]). These additional spikes are undesirable from an inference perspective, as they may be confused with the one contributed by the unknown parameter β^*.

The proportionality between parameter dimension d and sample size n in assumption (A4) is a natural scaling since the spectral estimator starts to be correlated with β^* in this regime.

2.2. Spectral estimator

The spectral estimator is defined as

$$\beta^{\text{spec}}(y, X) := v_1(D) \in \mathbb{S}^{d-1},$$

where $v_1(\cdot)$ denotes the principal eigenvector. We also define random variables

$$(\bar{G}, \bar{\varepsilon}) \sim \mathcal{N}\left(0, \frac{1}{\delta}\mathbb{E}[\bar{\Sigma}]\right) \otimes P_\varepsilon, \quad \bar{Y} = q(\bar{G}, \bar{\varepsilon}), \tag{2.2}$$

and an auxiliary function $\mathcal{F}_a : \mathbb{R} \to \mathbb{R}$ (for any $a > \sup \text{supp}(\mathcal{T}(\bar{Y}))$):

$$\mathcal{F}_a(\cdot) := \frac{\mathcal{T}(\cdot)}{a - \mathcal{T}(\cdot)}. \tag{2.3}$$

We make the following assumption on the preprocessing function:

(A5) $\mathcal{T}: \mathbb{R} \to \mathbb{R}$ is bounded and satisfies

$$\sup_{y \in \mathrm{supp}(\bar{Y})} \mathcal{T}(y) > 0. \tag{2.4}$$

Furthermore, $\mathcal{T}$ is pseudo-Lipschitz of finite order, i.e., there exist j and L such that

$$|\mathcal{T}(x) - \mathcal{T}(y)| \le L|x - y|\big(1 + |x|^{j-1} + |y|^{j-1}\big) \quad \text{for all } x, y.$$

The condition in (2.4) is rather mild: it is satisfied by the optimal preprocessing function (see Theorem 3.2), and it is also required by prior work for $\Sigma = I_d$ (see [56, 62]).

Finally, we single out two technical conditions that guarantee the well-posedness of the auxiliary quantities appearing in the statement of our main result, Theorem 3.1:

(A6) For any $x \ne 0$, let

$$\vartheta := \begin{cases} x \cdot (\sup \mathrm{supp}(\bar{\Sigma})), & x > 0, \\ x \cdot (\inf \mathrm{supp}(\bar{\Sigma})), & x < 0, \end{cases}$$

where we use $\mathrm{supp}(\bar{\Sigma})$ to denote the support of the density function of $\bar{\Sigma}$. Then for any $x \ne 0$, the random variable $\bar{\Sigma}$ satisfies

$$\lim_{\gamma \searrow \vartheta} \mathbb{E}\left[\frac{\bar{\Sigma}}{\gamma - x\bar{\Sigma}}\right] \overset{(a)}{=} \lim_{\gamma \searrow \vartheta} \mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma - x\bar{\Sigma})^2}\right] \overset{(b)}{=} \lim_{\gamma \searrow \vartheta} \mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma - x\bar{\Sigma})^2}\right] \overset{(c)}{=} \infty. \tag{2.5}$$

(A7) The function $\mathcal{T}$ satisfies

$$\lim_{a \searrow \sup \mathrm{supp}(\mathcal{T}(\bar{Y}))} \mathbb{E}\big[\mathcal{F}_a(\bar{Y})\big] \overset{(d)}{=} \lim_{a \searrow \sup \mathrm{supp}(\mathcal{T}(\bar{Y}))} \mathbb{E}\big[\bar{G}^2 \mathcal{F}_a(\bar{Y})\big] \overset{(e)}{=} \infty. \tag{2.6}$$

Assumptions (A6) and (A7) are mild and common in related work. In fact, assumption (A7) appeared in a similar form in [55, (A.5)], and it is common in the random matrix theory literature as well (see, e.g., [30, Assumption 6 "Thickness of the bulk edge", p. 129]). Assumption (A6) is similar to assumption (A7), but instead imposed on $\bar{\Sigma}$. Note that assumptions (A6) and (A7) do not a priori impose any dependence of $\mathcal{T}$ on q or $\bar{\Sigma}$. We remark that these two conditions can be removed, at the price of a slightly more involved definition of such auxiliary quantities; see Remark 3.3.

3. Main results

Our main contribution, Theorem 3.1, gives a precise asymptotic characterization of the overlap between the leading eigenvector of D and the unknown parameter, provided a criticality condition is satisfied. This condition ensures that D has a spectral gap in the high-dimensional limit. Theorem 3.1 also gives asymptotic formulas for the location of the right edge of the bulk and for the (right) outlier eigenvalue of D.

To state the results, we require some definitions. For $a \in (\sup \operatorname{supp}(\mathcal{T}(\bar{Y})), \infty)$, let

$$s(a) := \begin{cases} (\sup \operatorname{supp}(\bar{\Sigma})) \mathbb{E}[\mathcal{F}_a(\bar{Y})], & \mathbb{E}[\mathcal{F}_a(\bar{Y})] > 0, \\ (\inf \operatorname{supp}(\bar{\Sigma})) \mathbb{E}[\mathcal{F}_a(\bar{Y})], & \mathbb{E}[\mathcal{F}_a(\bar{Y})] < 0, \\ 0, & \mathbb{E}[\mathcal{F}_a(\bar{Y})] = 0. \end{cases} \tag{3.1}$$

Note that $s(a)$ also depends on $\bar{\Sigma}$ and $\mathcal{T}$. For $a > \sup \operatorname{supp}(\mathcal{T}(\bar{Y}))$, define the function

$$\varphi(a) = \frac{a}{\mathbb{E}[\bar{\Sigma}]} \mathbb{E}[\bar{G}^2 \mathcal{F}_a(\bar{Y})] \mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\Sigma}}\right], \quad \psi(a) = a\gamma(a), \tag{3.2}$$

where $\gamma(a)$ is an implicit function of a given by the unique solution in $(s(a), \infty)$ to

$$1 = \frac{1}{\delta} \mathbb{E}\left[\frac{\bar{\Sigma}}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\Sigma}}\right]. \tag{3.3}$$

To see existence and uniqueness of the solution, note that for any $a > \sup \operatorname{supp}(\mathcal{T}(\bar{Y}))$ such that $\mathbb{E}[\mathcal{F}_a(\bar{Y})] \neq 0$, $\frac{1}{\delta}\mathbb{E}[\bar{\Sigma}/(\gamma - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\Sigma})]$ is a strictly decreasing (since $\bar{\Sigma}$ is strictly positive) function of γ which approaches ∞ as $\gamma \searrow s(a)$ (see (a) in (2.5)) and approaches 0 as $\gamma \nearrow \infty$. If $\mathbb{E}[\mathcal{F}_a(\bar{Y})] = 0$, the solution $\gamma(a) = \frac{1}{\delta}\mathbb{E}[\bar{\Sigma}] > 0$ is obviously unique.

Next, using ψ and φ, we define two parameters a^*, a° that govern the validity of our spectral characterization. It can be shown (see Lemma E.3) that ψ is differentiable and has at least one critical point. Let $a^\circ > \sup \operatorname{supp}(\mathcal{T}(\bar{Y}))$ be the largest solution to
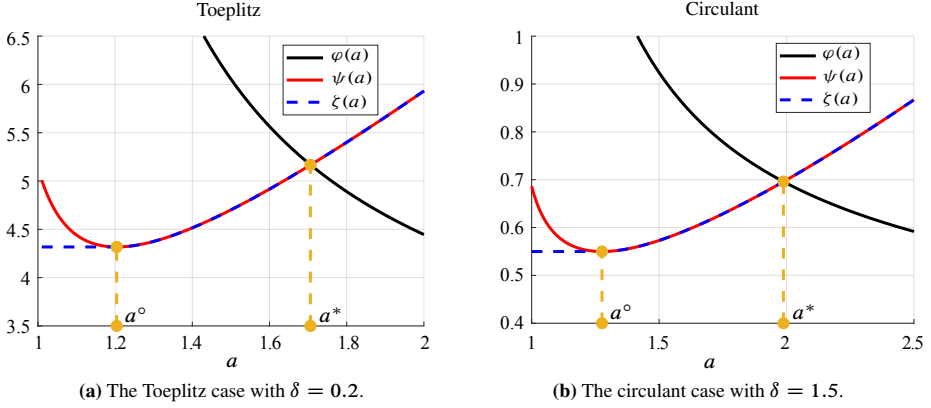
$$\psi'(a^\circ) = 0. \tag{3.4}$$

We then define $\zeta: (\sup \operatorname{supp}(\mathcal{T}(\bar{Y})), \infty) \to \mathbb{R}$ by flattening ψ to the left of a°:

$$\zeta(a) := \psi\big(\max\{a, a^\circ\}\big). \tag{3.5}$$

Finally, let a^* be the largest solution in $(\sup \operatorname{supp}(\mathcal{T}(\bar{Y})), \infty)$ to the following equation:

$$\zeta(a^*) = \varphi(a^*). \tag{3.6}$$

Proposition D.1 shows that such a solution must exist. The functions φ, ψ, ζ are plotted in Figure 1 for two examples of covariance matrix Σ.

(a) The Toeplitz case with $\delta = 0.2$.   (b) The circulant case with $\delta = 1.5$.

Figure 1. Plots of the functions $\varphi, \psi, \zeta : (\sup \mathrm{supp}(\mathcal{T}^*(\bar{Y})), \infty) \to \infty$ defined in (3.2) and (3.5) with $\mathcal{T}^*$ obtained by truncating the optimal preprocessing and $\bar{\bar{\Sigma}}$ given by the Toeplitz or circulant matrices (see Section 4.1.1 for details).

Then, the limits of the top two eigenvalues of D are given by

$$\lambda_1 := \zeta(a^*), \quad \lambda_2 := \zeta(a^\circ), \tag{3.7}$$

and the asymptotic overlap admits the following explicit expression:

$$\eta := \left(\frac{(1-w_2)\mathbb{E}\left[\frac{\bar{\bar{\Sigma}}}{\gamma(a^*)-\mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\bar{\Sigma}}}\right]^2}{(1-w_2)\mathbb{E}\left[\frac{\bar{\bar{\Sigma}}^2}{(\gamma(a^*)-\mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\bar{\Sigma}})^2}\right] + w_1\mathbb{E}\left[\frac{\bar{\bar{\Sigma}}}{(\gamma(a^*)-\mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\bar{\Sigma}})^2}\right]} \right)^{1/2}, \tag{3.8}$$

where the function $\mathcal{F}_{a^*}(\cdot)$ is defined in (2.3) and the ancillary parameters w_1, w_2 are given by:

$$w_1 := \frac{1}{\delta\mathbb{E}[\bar{\bar{\Sigma}}]}\mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\bar{\Sigma}}]}\bar{G}^2 - 1\right)\mathcal{F}_{a^*}(\bar{Y})^2\right]\mathbb{E}\left[\frac{\bar{\bar{\Sigma}}^2}{\gamma(a^*)-\mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\bar{\Sigma}}}\right]^2$$
$$+ \frac{1}{\delta}\mathbb{E}\left[\mathcal{F}_{a^*}(\bar{Y})^2\right]\mathbb{E}\left[\frac{\bar{\bar{\Sigma}}^3}{(\gamma(a^*)-\mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\bar{\Sigma}})^2}\right], \tag{3.9}$$

$$w_2 := \frac{1}{\delta}\mathbb{E}\left[\mathcal{F}_{a^*}(\bar{Y})^2\right]\mathbb{E}\left[\frac{\bar{\bar{\Sigma}}^2}{(\gamma(a^*)-\mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\bar{\Sigma}})^2}\right]. \tag{3.10}$$

We note that, given $a^* > a^\circ$, η is well defined as the fraction under the square root is strictly positive. This is because (i) all three expectations in (3.8) are positive ($\bar{\bar{\Sigma}} > 0$ in assumption (A2) and $\gamma(a^*) > s(a^*)$); (ii) $w_1 > 0$ (see Proposition G.1); and (iii) $1 - w_2 > 0$ if $a^* > a^\circ$ (see item (3) of Proposition D.6).

We are now ready to state our main result, whose proof is given in Section 5, with several details deferred to Section A.

Theorem 3.1 (Performance characterization of spectral estimator). *Consider the setting of Section 2 and let assumptions (A1) to (A7) hold. Suppose $a^* > a^\circ$. Then, the top two eigenvalues $\lambda_1(D)$, $\lambda_2(D)$ of D satisfy*[3]

$$\operatorname*{p-lim}_{d\to\infty} \lambda_1(D) = \lambda_1, \quad \lim_{d\to\infty} \lambda_2(D) = \lambda_2 \quad \text{almost surely,} \tag{3.11}$$

and $\lambda_1 > \lambda_2$, where p-lim *denotes the limit in probability. Furthermore, the limiting overlap between the top eigenvector $v_1(D)$ and β^* equals*

$$\operatorname*{p-lim}_{d\to\infty} \frac{|\langle v_1(D), \beta^* \rangle|}{\|\beta^*\|_2} = \eta > 0. \tag{3.12}$$

Remark 3.1 (Uniqueness of a^*). Recall that the parameter a^* is the largest solution in $(\sup \operatorname{supp}(\mathcal{T}(\bar{Y})), \infty)$ to (3.6). With additional assumptions, we can show that (3.6) admits a unique solution; see Proposition D.3 for details. We expect that the additional assumptions can be removed and the solution to (3.6) in $(\sup \operatorname{supp}(\mathcal{T}(\bar{Y})), \infty)$ always exists and is unique.

Remark 3.2 (Consistency with isotropic covariance). We note that, by setting $\Sigma = I_d$, we recover the existing result on i.i.d. Gaussian designs (i.e., [62, Lemma 2]).

Remark 3.3 (Removing assumptions (A6) and (A7)). We first note that assumption (A6) requires $\operatorname{law}(\bar{\Sigma})$ to have sufficiently slow decay on both the left and right edges, and assumption (A7) requires such behavior on the right edge of $\operatorname{law}(\mathcal{T}(\bar{Y}))$. However, both assumptions can be removed at the cost of a vanishing perturbation of $\bar{\Sigma}$, $\mathcal{T}$ around their edges in the definitions of λ_1, λ_2, η in (3.7) and (3.8). The perturbed quantities, denoted by λ_1', λ_2', η', are guaranteed to satisfy both assumptions. Hence, Theorem 3.1 ensures that the high-dimensional limits of the top two eigenvalues and of the overlap for the perturbed matrix D' are given by λ_1', λ_2', η', respectively. An application of the Davis–Kahan theorem [26] shows that, as the perturbation vanishes, the top two eigenvalues and overlap obtained with D' coincide with those given by the unperturbed matrix D. Furthermore, since λ_1', λ_2', η' are continuous with respect to the perturbation, their limits as the perturbation vanishes exist. Therefore, the latter limits must equal the high-dimensional limits of the top two eigenvalues and overlap given by the original D. The formal argument is deferred to Section C, and by a similar argument, assumptions (A6) and (A7) in Theorem 3.2 below can be removed as well.

---

[3]For a symmetric matrix $M \in \mathbb{R}^{p \times p}$, we write its (real) eigenvalues as $\lambda_1(M) \geq \cdots \geq \lambda_p(M)$ and the associated eigenvectors (normalized to have unit ℓ_2-norm) as $v_1(M), \ldots, v_p(M) \in \mathbb{S}^{p-1}$.

Remark 3.4 (Phase transition). Our characterization of the outlier eigenvalue and the overlap is valid given an explicit and checkable condition $a^* > a^\circ$ not depending on the data (y, X). Informally, it guarantees that the aspect ratio δ exceeds a certain threshold which leads to a spike in D. We conjecture that this condition is in fact necessary, in the sense that otherwise the spectral estimator fails to achieve a positive limiting overlap and the top eigenvalue sticks to the bulk of the spectrum of D. It is easy to verify that $\lambda_1 = \lambda_2$ and $\eta = 0$ precisely when $a^* = a^\circ$, indicating a continuous phase transition at the conjectured threshold.

Remark 3.5 (Non-i.i.d. prior). We expect that all results in the paper can be extended to the more general setting in which β^* may be asymptotically aligned with the eigenvectors of Σ. We describe the required modifications below. Suppose that $\beta^* \in \mathbb{R}^d$ is such that p-$\lim_{d \to \infty} d^{-1} \|\beta^*\|_2^2 = 1$ and the following empirical probability measure admits a weak limit in probability:

$$\varrho := \sum_{i=1}^{d} \frac{\langle \beta^*, v_i(\Sigma) \rangle^2}{\|\beta^*\|_2^2} \delta_{\lambda_i(\Sigma)} \xrightarrow[d \to \infty]{\mathbb{P}} \overline{\varrho}.$$

The measure ϱ records the alignment between β^* and each of the eigenvectors of Σ. Note that in the special case where β^* has i.i.d. entries (with mean 0 and variance 1) considered in the paper, $\overline{\varrho}$ equals the law of $\overline{\Sigma}$. This can be seen by examining the convergence of the Stieltjes transform of ϱ. In the state evolution analysis, whenever the limit of $d^{-1} \mathbb{E}[\langle \beta^*, f(\Sigma)\beta^* \rangle]$ (for some $f : \mathbb{R} \to \mathbb{R}$ that applies to Σ according to functional calculus) needs to be computed, we would write

$$\lim_{d \to \infty} \frac{1}{d} \mathbb{E}\big[\langle \beta^*, f(\Sigma)\beta^* \rangle\big] = \lim_{d \to \infty} \mathbb{E}\left[\frac{\|\beta^*\|_2^2}{d} \sum_{i=1}^{d} \frac{\langle \beta^*, v_i(\Sigma) \rangle^2}{\|\beta^*\|_2^2} f(\lambda_i(\Sigma))\right]$$

$$= \lim_{d \to \infty} \mathbb{E}\left[\frac{\|\beta^*\|_2^2}{d} \int f(\lambda)\varrho(\mathrm{d}\lambda)\right]$$

$$= \int f(\lambda)\overline{\varrho}(\mathrm{d}\lambda).$$

We expect that such modifications can allow us to obtain analogous results for general β^* potentially correlated with eigenvectors of Σ.

Remark 3.6 (Spectrally initialized AMP). Whenever δ is large enough so that our spectral estimator becomes effective, one can analyze spectrally initialized Bayes-AMP by running AMP with linear denoisers as designed in the paper for a large (but constant with respect to n, d) number of steps and then transitioning to Bayes-optimal denoisers. This corresponds to the strategy pursued in [66]. However, we note that Bayes-AMP in general requires the knowledge of Σ which is not assumed to be available in our paper for the design of spectral estimators.

3.1. Optimal spectral methods for general Gaussian designs

Theorem 3.1 holds for an arbitrary function $\mathcal{T}$ subject to mild regularity conditions. This enables the optimization of $\mathcal{T}$ to minimize the spectral threshold, i.e., the smallest δ such that $a^* > a^\circ$. The result on the optimization of the preprocessing function is stated below and proved in Section B.

Theorem 3.2 (Optimal spectral threshold). *Consider the setting of Section 2, let assumptions (A1) to (A4), and (A6) hold, and let $\mathfrak{T}$ be the set of functions $\mathcal{T} \colon \mathbb{R} \to \mathbb{R}$ satisfying assumptions (A5) and (A7). Then the following two statements hold.*

(1) *There exists $\mathcal{T} \in \mathfrak{T}$ such that $a^* > a^\circ$ holds if*

$$\delta > \Delta(\delta) := \frac{\mathbb{E}[\bar{\Sigma}]^2}{\mathbb{E}[\bar{\Sigma}^2]} \left(\int_{\mathrm{supp}(\bar{Y})} \frac{\mathbb{E}\left[p(y \mid \bar{G}) \left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]} \bar{G}^2 - 1 \right) \right]^2}{\mathbb{E}[p(y \mid \bar{G})]} \, \mathrm{d}y \right)^{-1}, \quad (3.13)$$

with $p(\bar{Y} \mid \bar{G})$ the conditional density of $\bar{Y} = q(\bar{G}, \bar{\varepsilon})$ given $\bar{G}$, determined via the joint distribution in (2.2). In this case, if

$$\mathcal{T}^*(y) = 1 - \left(\sqrt{\frac{\Delta(\delta)}{\delta}} \frac{\mathbb{E}\left[p(y \mid \bar{G}) \left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]} \bar{G}^2 \right) \right]}{\mathbb{E}[p(y \mid \bar{G})]} + 1 - \sqrt{\frac{\Delta(\delta)}{\delta}} \right)^{-1} \quad (3.14)$$

is pseudo-Lipschitz of finite order, then the spectral estimator using the pre-processing function $\mathcal{T}^*$ achieves strictly positive limiting overlap.

(2) *Conversely, suppose that the function φ defined in (3.2) is strictly decreasing for every $\mathcal{T} \in \mathfrak{T}$. If there exists $\mathcal{T} \in \mathfrak{T}$ such that $a^* > a^\circ$, then δ satisfies (3.13).*

Remark 3.7 (Mild dependence of $\mathcal{T}^*$ on $\bar{\Sigma}$). The optimal function $\mathcal{T}^*$ in (3.14) depends on $\bar{\Sigma}$ only through its first moment, or equivalently it depends on Σ only through its normalized trace. We highlight that approximating $\frac{1}{d} \mathrm{Tr}(\Sigma)$ from the data is significantly easier than approximating the whole matrix Σ. In fact, $\frac{1}{d} \mathrm{Tr}(\Sigma)$ can be estimated consistently via the plugin estimator $\frac{1}{d} \mathrm{Tr}(X^\top X)$. Specifically, achieving a root mean square error of ς only requires $n = \mathcal{O}(\varsigma^{-2})$, which is trivially satisfied by assumption (A4). In contrast, the sample complexity needed to estimate Σ with sufficient accuracy may be larger than that required by the spectral estimator itself. Specifically, achieving an error of ς in spectral norm for the estimation of Σ via the sufficient statistic $X^\top X$ requires $n = \Theta(d\varsigma^{-2})$; see [74, Exercise VI.15], [92, Section 24.2]. Note that, to estimate Σ, n scales linearly with d and the proportionality constant may be larger than the critical value of δ in the right-hand side of (3.13); instead, to estimate $\frac{1}{d} \mathrm{Tr}(\Sigma)$, n does not depend on d and, hence, the estimate is consistent for all $\delta > 0$.

Remark 3.8 (Sufficient condition for $\mathcal{T}^*$ being pseudo-Lipschitz). The assumption in Theorem 3.2 that $\mathcal{T}^*$ is pseudo-Lipschitz of finite order is satisfied by models that contain an additive component of Gaussian noise (regardless of the variance of the Gaussian noise). This requirement is mild, and common in the related literature; see, e.g., [4]. Specifically, consider the GLM $y = \widetilde{q}(X\beta^*, \varepsilon') + \varepsilon''$, where $\widetilde{q}(X\beta^*, \varepsilon')$ satisfies assumptions (A1) to (A4), and (A6) and is independent of $\varepsilon'' \sim \mathcal{N}(0_n, \sigma^2 I_n)$ (for some $\sigma > 0$). Then, one can verify that

$$\frac{\mathbb{E}\big[p(y \mid \bar{G})\big(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2\big)\big]}{\mathbb{E}[p(y \mid \bar{G})]},$$

and hence $\mathcal{T}^*(y)$, is pseudo-Lipschitz of finite order.

Remark 3.9 (Monotonicity of φ). The second part of Theorem 3.2 assumes the monotonicity of φ. One readily checks that this holds when $\bar{\Sigma} = 1$ (i.e., $\Sigma = I_d$). Furthermore, in Section D.1, we prove that φ is strictly decreasing for non-negative $\mathcal{T}$ (Proposition D.2) and give numerical evidence that the same result holds for general $\mathcal{T}$ (Remark D.1).

3.2. Optimal spectral methods for rotationally invariant designs

The inequality (3.13) can be interpreted as giving the optimal spectral threshold, i.e., the minimal δ above which positive overlap is achievable by some spectral estimator. Furthermore, this threshold is attained by $\mathcal{T}^*$ in (3.14). As δ gets close to the spectral threshold $\Delta(\delta)$, $\mathcal{T}^*$ approaches the following function (obtained by replacing $\sqrt{\Delta(\delta)/\delta}$ in $\mathcal{T}^*$ with 1):

$$\mathcal{T}^\star(y) = 1 - \frac{\mathbb{E}[p(y \mid \bar{G})]}{\mathbb{E}\big[p(y \mid \bar{G})\big(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2\big)\big]}.$$

When $\Sigma = I_d$, $\mathcal{T}^\star$ minimizes the spectral threshold [62] and maximizes the overlap for any δ above that threshold [56]. Supported by evidence from statistical physics, it is conjectured in [59, Conjecture 2] that the optimality to hold for the more general ensemble of right rotationally invariant designs. Although our design X is only left rotationally invariant, if the unknown parameter is Gaussian ($\beta^* \sim \mathcal{N}(0_d, I_d)$) or uniform on the sphere ($\beta^* \sim \text{Unif}(\sqrt{d}\, \mathbb{S}^{d-1})$), the model in (1.1) is equivalent to one with a design that is also right rotationally invariant. Therefore, *Theorem* 3.2 *proves* [59, Conjecture 2] *for a class of spectral distributions of X* – specifically, those given by the multiplicative free convolution of the Marchenko–Pastur law with a measure compactly supported on $(0, \infty)$. Formally, with the following two assumptions in place of assumptions (A1) and (A2), Theorem 3.2 implies Corollary 3.3:

(A8) $\beta^* \sim \mathrm{Unif}(\sqrt{d}\,\mathbb{S}^{d-1})$ or $\beta^* \sim \mathcal{N}(0_d, I_d)$.

(A9) $X = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top \in \mathbb{R}^{n \times d}$ can be written as $X = BQ^\top$, with the rows of $B \in \mathbb{R}^{n \times d}$ satisfying assumption (A2) and $Q \sim \mathrm{Haar}(\mathbb{O}(d))$ independent of everything else, where $\mathbb{O}(d)$ is the orthogonal group in dimension d.

Corollary 3.3. *Consider the setting of Section* 2 *and let assumptions* (A3), (A4), (A6), (A8), *and* (A9) *hold. Then, the conclusions of Theorem* 3.2 *hold.*

Proof. By assumption (A9), $X = \widetilde{B}\Sigma^{1/2}Q^\top$, where $\widetilde{B} \in \mathbb{R}^{n \times d}$ has i.i.d. $\mathcal{N}(0, 1/n)$ entries and $\Sigma \in \mathbb{R}^{d \times d}$ is a covariance matrix satisfying assumption (A2). Let

$$
\begin{aligned}
D &= X^\top \mathrm{diag}\big(\mathcal{T}(q(X\beta^*, \varepsilon))\big) X \\
&= Q\Sigma^{1/2}\widetilde{B}^\top \mathrm{diag}\big(\mathcal{T}(q(\widetilde{B}\Sigma^{1/2}Q^\top\beta^*, \varepsilon))\big)\widetilde{B}\Sigma^{1/2}Q^\top, \\
\widehat{D} &= \Sigma^{1/2}\widetilde{B}^\top \mathrm{diag}\big(\mathcal{T}(q(\widetilde{B}\Sigma^{1/2}Q^\top\beta^*, \varepsilon))\big)\widetilde{B}\Sigma^{1/2}, \\
\widetilde{D} &= \Sigma^{1/2}\widetilde{B}^\top \mathrm{diag}\big(\mathcal{T}(q(\widetilde{B}\Sigma^{1/2}\beta^*, \varepsilon))\big)\widetilde{B}\Sigma^{1/2}.
\end{aligned}
$$

Then, we have

$$
\frac{|\langle v_1(D), \beta^*\rangle|}{\|\beta^*\|_2} = \frac{|\langle Qv_1(\widehat{D}), \beta^*\rangle|}{\|\beta^*\|_2} = \frac{|\langle v_1(\widehat{D}), Q^\top\beta^*\rangle|}{\|Q^\top\beta^*\|_2} \overset{\mathrm{d}}{=} \frac{|\langle v_1(\widetilde{D}), \beta^*\rangle|}{\|\beta^*\|_2}. \quad (3.15)
$$

The first equality uses that, if (λ, v) is an eigenpair of D, then (λ, Qv) is an eigenpair of $QDQ^\top$ for $Q \in \mathbb{O}(d)$. The second equality holds as Q is orthogonal. The third passage follows since by assumption (A8), $\beta^* \overset{\mathrm{d}}{=} Q^\top\beta^*$ for $Q \in \mathrm{Haar}(\mathbb{O}(d))$ independent of β^*. Now Theorem 3.2 applies to the rightmost side of (3.15), which completes the proof. ∎

4. Numerical experiments

We consider noiseless phase retrieval ($y_i = |\langle x_i, \beta^*\rangle|$) in Section 4.1 and Poisson regression ($y_i \sim \mathrm{Pois}(\langle x_i, \beta^*\rangle^2)$) in Section 4.2, and evaluate the performance of the spectral estimator with different preprocessing functions. In all plots, "sim" and "thy" in legends denote simulation results and theoretical predictions, respectively.

4.1. Phase retrieval

4.1.1. Synthetic data. For all the synthetic experiments, we take the parameter $\beta^* \sim \mathrm{Unif}(\sqrt{d}\,\mathbb{S}^{d-1})$ and $d = 2000$. We plot the overlap between the spectral estimator and β^*, as a function of the aspect ratio δ. Each value is computed from 10 i.i.d.

trials, the error bar is at 1 standard deviation, and the corresponding theoretical predictions are continuous lines with the same color. We consider three types of covariance matrix Σ:

(i) Toeplitz covariance, $\Sigma_{i,j} = \rho^{|i-j|}$ for $1 \leq i, j \leq d$ with $\rho = 0.9$, as considered in [97, Section 4] and [45, Section 5.3].

(ii) Circulant covariance, $\Sigma_{i,j} = c_0$ for $i = j$, $\Sigma_{i,j} = c_1$ for $i + 1 \leq j \leq i + \ell$ and $i + d - \ell \leq j \leq i + d - 1$, $\Sigma_{i,j} = 0$ otherwise, with $c_0 = 1, c_1 = 0.1, \ell = 17$, as considered in [44, Section F] and [43, Section 5.1].

(iii) Identity covariance, $\Sigma = I_d$.

We compare spectral estimators using different preprocessing functions:

(i) The optimal choice in (3.14) with truncation, i.e.,

$$\mathcal{T}^*(y) = \max\{1 - \mathbb{E}[\bar{\Sigma}]/(\delta y^2), -K_*\},$$

with $K_* = 10$. The truncation ensures that the preprocessing is bounded as required by our theory and, by taking K_* sufficiently large, it does not affect performance.

(ii) The trimming scheme [21], i.e.,

$$\mathcal{T}^{\text{trim}}(y) = \delta y^2/\mathbb{E}[\bar{\Sigma}]\mathbb{1}\{\sqrt{\delta}|y|/\sqrt{\mathbb{E}[\bar{\Sigma}]} \leq K_{\text{trim}}\},$$

with $K_{\text{trim}} = \sqrt{7}$.

(iii) The subset scheme [89], i.e.,

$$\mathcal{T}^{\text{subset}}(y) = \mathbb{1}\{\sqrt{\delta}|y|/\sqrt{\mathbb{E}[\bar{\Sigma}]} \geq K_{\text{subset}}\},$$

with $K_{\text{subset}} = \sqrt{2}$. The values of both K_{trim} and K_{subset} are taken from [62, Section 7.1] where they are optimized to yield the smallest spectral threshold for $\Sigma = I_d$.

(iv) The identity function with truncation, i.e.,

$$\mathcal{T}^{\text{id}}(y) := \min\{\max\{\sqrt{\delta}y/\sqrt{\mathbb{E}[\bar{\Sigma}]}, -K_{\text{id}}\}, K_{\text{id}}\},$$

with K_{id} equal to 3.5 and 3 for circulant and Toeplitz covariances, respectively. Empirically, the performance under these choices of K_{id} does not differ much from avoiding the truncation, i.e., $K_{\text{id}} = \infty$.

We also compare the performance with a *whitened* spectral estimator, which requires knowledge of the covariance Σ. The whitened spectral estimator is given by

$$\beta_{\triangle}^{\text{spec}} := \Sigma^{-1/2}v_1(D_{\triangle}), \tag{4.1}$$
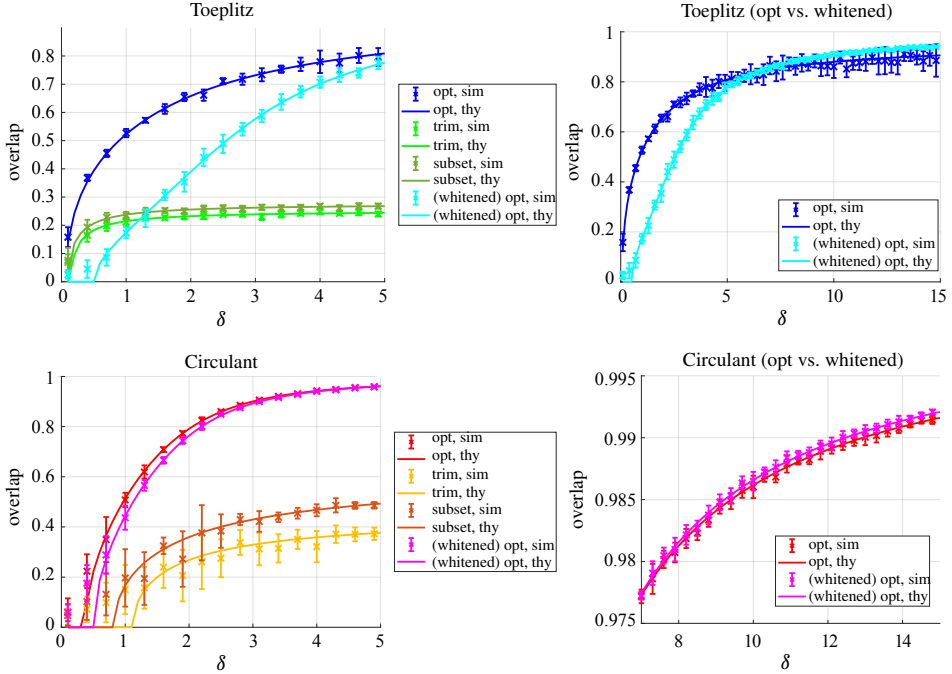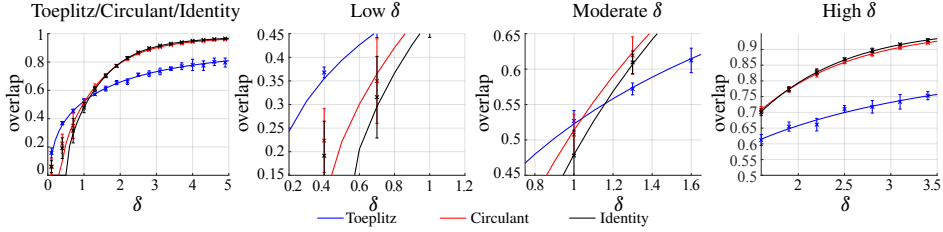
Figure 2. Overlap of spectral estimators with different preprocessing functions for noiseless phase retrieval when the covariate vectors are independent zero-mean Gaussians with Toeplitz (top row) and circulant (bottom row) covariance.

where $D_\triangle := (X\Sigma^{-1/2})^\top \operatorname{diag}(\mathcal{T}(y))(X\Sigma^{-1/2})$. This estimator uses Σ to whiten X and computes the principal eigenvector of $D_\triangle$ obtained via the decorrelated covariates $X\Sigma^{-1/2}$. As the eigenvector can be thought of as an estimate of $\Sigma^{1/2}\beta^*$, it is left-multiplied by $\Sigma^{-1/2}$ to produce an estimate of β^*. Formal results and proofs on $\beta_\triangle^{\mathrm{spec}}$ are deferred to Section F.

Figure 2 shows that our proposed optimal spectral estimator significantly outperforms the trimming/subset schemes for both Toeplitz (top) and circular (bottom) covariances. Furthermore, in a large interval of δ, the performance of the whitened spectral estimator in (4.1) (which requires Σ) is significantly worse than that of the standard spectral estimator (which does not require Σ), even though optimal preprocessing functions are employed for both.

In Figure 3, the plots for Toeplitz, circulant and identity covariance are superimposed. An interesting observation is that there is no universally best covariance structure, even if the optimal preprocessing function with respect to the corresponding covariance is adopted.
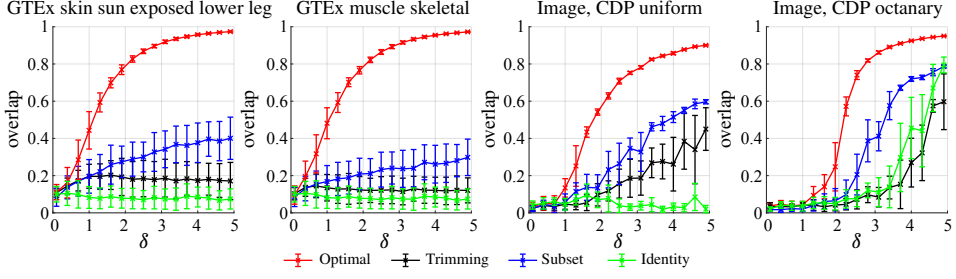
Figure 3. Overlap of spectral estimators with optimal preprocessing function for noise-less phase retrieval when the covariate vectors are independent zero-mean Gaussians with Toeplitz/circulant/identity covariance. The right three panels respectively zoom into regimes where δ takes low, moderate and high values to demonstrate that in this particular setting, any one of the three types of covariance structures can attain the highest overlap.

4.1.2. Real data. We also demonstrate the advantage of the optimal preprocessing given by our theory for data sets popular in quantitative genetics and computational imaging.

Specifically, the design matrices for the first two plots of Figure 4 are obtained from two GTEx data sets "skin sun exposed lower leg" (56200×701) and "muscle skeletal" (56200×803) [54]. These matrices record gene counts and therefore contain non-negative entries. We preprocess them as follows: (i) remove all-0 rows, (ii) build a matrix by sequentially including each row only if it has an overlap smaller than 0.3 with all existing rows, and (iii) center and normalize each row such that it has zero mean and unit variance. All these operations are typical in genetic studies; see, e.g., the widely used tool set PLINK [19]. The unknown parameter vector is given by $\beta^* \sim \text{Unif}(\sqrt{d}\, \mathbb{S}^{d-1})$ for $d \in \{701, 803\}$. For each δ, the design matrix is formed by the first $\lfloor d\delta \rfloor$ rows of the above preprocessed matrix. The value of overlap for each δ is computed from 100 i.i.d. trials where the randomness is only over β^*, and the error bar is reported at 1 standard deviation. The truncation levels for different preprocessing functions are chosen as follows: for $\mathcal{T}^*$, we set $K_* = 100$; for $\mathcal{T}^{\text{trim}}$ and $\mathcal{T}^{\text{subset}}$, for each δ, we choose K_{trim} and K_{subset} in $\{0.25i : 1 \leq i \leq 40\}$ to maximize the respective overlaps (averaged over 100 trials); for $\mathcal{T}^{\text{id}}$, we do not truncate, i.e., $K_{\text{id}} = \infty$. Despite the advantage due to the adaptive choice of the truncation level for the trimming/subset scheme, the preprocessing we propose still performs vastly better than all alternatives.

The design matrices for the last two plots of Figure 4 follow a coded diffraction pattern [13], i.e., X is obtained by stacking in its rows the matrices FD_1S, $FD_2S, \ldots, FD_\delta S$. Here, $\delta \in \mathbb{Z}_{\geq 1}$, $F \in \mathbb{R}^{d \times d}$ is a Discrete Fourier Transform matrix, $S \in \mathbb{R}^{d \times d}$ is diagonal containing i.i.d. uniformly random signs, and $D_1, D_2, \ldots, D_\delta \in \mathbb{C}^{d \times d}$ are diagonal with elements following one of these two distributions:

Figure 4. Overlap of spectral estimators for noiseless phase retrieval when the design matrix is obtained from two Genotype-Tissue Expression (GTEx) data sets (first two plots) and two coded diffraction patterns (CDP) (last two plots).

(i)   uniform modulation, $(D_\ell)_{i,i} \overset{\text{i.i.d.}}{\sim} \text{Unif}([-10, 10])$, and

(ii) octanary modulation [13, (1.9)], $(D_\ell)_{i,i} \overset{\text{i.i.d.}}{\sim} \text{law}(\bar{D})$ with $\bar{D} = \bar{D}_1 \bar{D}_2$,

$$\text{law}(\bar{D}_1) = \frac{1}{4}(\delta_1 + \delta_{-1} + \delta_{-i} + \delta_i) \quad \text{and} \quad \text{law}(\bar{D}_2) = \frac{4}{5}\delta_{1/\sqrt{2}} + \frac{1}{5}\delta_{\sqrt{3}}.$$

For fractional $\delta \in (0, \infty)$, we first construct a matrix of size $\lceil \delta \rceil d \times d$ and then randomly subsample $\lfloor \delta d \rfloor - \lfloor \delta \rfloor d$ rows from the last block $FD_{\lceil \delta \rceil} S$ to obtain a design matrix of size $\lfloor \delta d \rfloor \times d$.

The parameter β^* in the last two plots of Figure 4 is a 75×64 RGB image of the painting "Girl with a Pearl Earring". The 3 color bands give 3 matrices in $[0, 256]^{75 \times 64}$. The parameter vectors $\beta_R^*, \beta_G^*, \beta_B^* \in \mathbb{S}^{d-1}$ (with $d = 75 \times 64 = 4800$) are then obtained by vectorizing, centering, and normalizing each of these matrices. For each δ, we have 5 i.i.d. trials where the randomness is only over X. In each trial, we compute 3 spectral estimators using the same X and observations $y_R, y_G, y_B \in \mathbb{R}^n$ generated from $\beta_R^*, \beta_G^*, \beta_B^*$ respectively. We report the mean of $5 \times 3 = 15$ overlaps for each δ with error bar at 1 standard deviation. The truncation levels for different preprocessing functions are $K_* = 10$, $K_{\text{trim}} = \sqrt{7}$, $K_{\text{subset}} = \sqrt{2}$, $K_{\text{id}} = \infty$. For all data sets, our proposed preprocessing (optimal in red) outperforms previous heuristic choices (trimming [21] in black, subset [89] in blue, and identity in green).

4.2. Poisson regression

We also consider the Poisson regression model $y_i \sim \text{Pois}(\langle x_i, \beta^* \rangle^2)$, where $\beta^* \sim \text{Unif}(\sqrt{d}\,\mathbb{S}^{d-1})$ and $d = 2000$. The covariance matrix Σ is taken to be Toeplitz or circulant with the same parameters as in Section 4.1.1. We again consider 3 preprocessing functions:

(i)   the optimal one $\mathcal{T}^*(y) = (y - \mathbb{E}[\bar{\Sigma}]/\delta)/(y + 1/2)$,
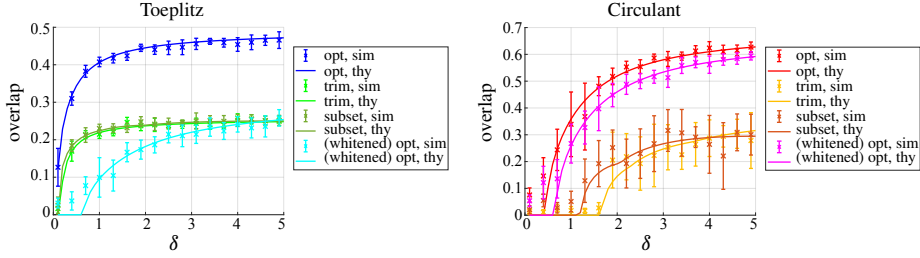
Figure 5. Overlap of spectral estimators with different preprocessing functions for Poisson regression with correlated Gaussian design with Toeplitz (left panel) and circulant (right panel) covariance.

(ii)   the trimming function $\mathcal{T}^{\mathrm{trim}}(y) = y\mathbb{1}\{|y| \leq K_{\mathrm{trim},\delta}\}$, and

(iii)   the subset function $\mathcal{T}^{\mathrm{subset}}(y) = \mathbb{1}\{|y| \geq K_{\mathrm{subset},\delta}\}$.

For each δ, $K_{\mathrm{trim},\delta}$, $K_{\mathrm{subset},\delta}$ are optimized over $[0.5, 50]$, $[0.5, 20]$, respectively, so as to maximize the overlap. Note that since y_i is $\mathbb{Z}_{\geq 0}$-valued, it suffices to consider $K_{\mathrm{trim},\delta}$, $K_{\mathrm{subset},\delta}$ of the form $K + 1/2$ for an integer K. The numerical results are shown in Figure 5.

5.   Proof of Theorem 3.1

5.1.   Overview of the argument

The outlier location and asymptotic overlap in Theorem 3.1 are derived using a variant of AMP for GLMs, known as *generalized approximate message passing* (GAMP) ([76], [40, Section 4]). An instance of GAMP is specified by two sequences of denoising functions, $(g_t)_{t \geq 0}$ and $(f_{t+1})_{t \geq 0}$. Starting with initialization $\widetilde{u}^{-1} = 0_n \in \mathbb{R}^n$ and some $\widetilde{v}^0 \in \mathbb{R}^d$, for $t \geq 0$ the GAMP iterates are computed as:

$$
\begin{aligned}
u^t &= \widetilde{X}\widetilde{v}^t - b_t\widetilde{u}^{t-1}, \quad \widetilde{u}^t = g_t(u^t; y), \\
c_t &= \frac{1}{n}\operatorname{div} g_t(u^t; y) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial g_t(u^t; y)_i}{\partial u_i^t}, \\
v^{t+1} &= \widetilde{X}^\top\widetilde{u}^t - c_t\widetilde{v}^t, \quad \widetilde{v}^{t+1} = f_{t+1}(v^{t+1}), \\
b_{t+1} &= \frac{1}{n}\operatorname{div} f_{t+1}(v^{t+1}) = \frac{1}{n}\sum_{i=1}^{d}\frac{\partial f_{t+1}(v^{t+1})_i}{\partial v_i^{t+1}},
\end{aligned}
\tag{5.1}
$$

where we recall $\widetilde{X} = X\Sigma^{-1/2}$. To handle $\Sigma \neq I_d$, the denoising functions $g_t \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ and $f_{t+1} \colon \mathbb{R}^d \to \mathbb{R}^d$ need to be *non-separable*, i.e., they cannot be decomposed in terms of functions acting component-wise on the vector inputs.

AMP algorithms come with an associated deterministic scalar recursion called *state evolution* which describes the limiting distribution (as $d \to \infty$) of the AMP iterates $u^t \in \mathbb{R}^n$ and $v^{t+1} \in \mathbb{R}^d$ using a collection of Gaussian vectors. The covariance structure of these Gaussians admits a succinct representation which can be recursively tracked via the state evolution. The state evolution result for GAMP with non-separable denoisers is not immediately available – we prove it by reducing such a GAMP to a general family of abstract AMP algorithms introduced in [42] for which a state evolution has been established. This is detailed in Section 5.2. We note that state evolution results for an abstract AMP similar to those in [42] can also be found in [81].

The key idea is to design a GAMP algorithm that simulates the power iteration $v^{t+1} = Dv^t / \|Dv^t\|_2$, via a careful choice of denoising functions g_t and f_{t+1}. To this end, we set

$$g_t(u^t; y) = Fu^t, \quad t \geq 0,$$

where $F = \operatorname{diag}(\mathcal{F}(y)) \in \mathbb{R}^{n \times n}$, and the functions $\mathcal{F}, (f_{t+1})_{t \geq 0}$ are specified later. With this choice for g_t, we have

$$c_t = \frac{1}{n} \sum_{i=1}^{n} \mathcal{F}(y_i) \xrightarrow{n \to \infty} \mathbb{E}[\mathcal{F}(\overline{Y})] =: c, \quad t \geq 0,$$

where we recall that $\overline{Y}$ is defined in (2.2). Thus, the GAMP iteration, with c_t replaced with its high-dimensional limit, becomes

$$u^t = \widetilde{X} f_t(v^t) - b_t Fu^{t-1}, \quad v^{t+1} = \widetilde{X}^\top Fu^t - cf_t(v^t).$$

We show in Section 5.5 that $u^t, v^{t+1}, b_t, f_{t+1}$ converge in probability as $t \to \infty$, i.e., there exist $u \in \mathbb{R}^n$, $v \in \mathbb{R}^d$, $b \in \mathbb{R}$ and $f \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ such that

$$\lim_{t \to \infty} \lim_{n \to \infty} \frac{1}{\sqrt{n}} \|u^t - u\|_2 = 0, \quad \lim_{t \to \infty} \lim_{d \to \infty} \frac{1}{\sqrt{d}} \|v^{t+1} - v\|_2 = 0,$$

$$\lim_{t \to \infty} \lim_{d \to \infty} |b_t - b| = 0, \quad \lim_{t \to \infty} \lim_{d \to \infty} \frac{1}{\sqrt{d}} \|f_{t+1}(v^{t+1}) - f(v)\|_2 = 0.$$

Thus, we obtain

$$u = \widetilde{X} f(v) - bFu, \quad v = \widetilde{X}^\top Fu - cf(v).$$

The first equation for u implies

$$u = (I_n + bF)^{-1} \widetilde{X} f(v).$$

Substituting this into the equation for v and multiplying both sides by $\Sigma^{1/2}$, we have

$$\Sigma^{1/2}\big(v + cf(v)\big) = \Sigma^{1/2}\widetilde{X}^\top F(I_n + bF)^{-1}\widetilde{X}\Sigma^{1/2}\Sigma^{-1/2}f(v). \qquad (5.2)$$

At this point, we consider the following choice of $\mathcal{F}$ and f:

$$\mathcal{F}(\cdot) = \frac{\mathcal{T}(\cdot)}{a - b\mathcal{T}(\cdot)}, \quad f(v) = (\gamma I_d - c\Sigma)^{-1}\Sigma v$$

for some $a, \gamma \in \mathbb{R}$ to be specified. Then, (5.2) becomes

$$\Sigma^{-1/2}f(v) = \frac{1}{a\gamma}\Sigma^{1/2}\widetilde{X}^\top T\widetilde{X}\Sigma^{1/2}\Sigma^{-1/2}f(v) = \frac{1}{a\gamma}D\Sigma^{-1/2}f(v),$$

which is an eigenequation of D with eigenvalue $a\gamma =: \lambda_1$ and eigenvector (possibly scaled by a constant) $\Sigma^{-1/2}f(v) = \Sigma^{-1/2}(\gamma I_d - c\Sigma)^{-1}\Sigma v$. Assuming a spectral gap, we expect that λ_1 equals the limiting value of $\lambda_1(D)$ and $\Sigma^{-1/2}f(v)$ is asymptotically aligned with $v_1(D)$.

It remains to pick a, γ which are in principle free parameters. Our choice is motivated by the fixed points of state evolution characterized in Section 5.3, and it simplifies the derivations. Specifically, the limiting Onsager coefficient is given by

$$b = \frac{1}{n}\sum_{i=1}^d \frac{\partial f(v)_i}{\partial v_i} = \frac{1}{n}\sum_{i=1}^d \big((\gamma I_d - c\Sigma)^{-1}\Sigma\big)_{i,i} \xrightarrow{n\to\infty} \frac{1}{\delta}\mathbb{E}\left[\frac{\overline{\Sigma}}{\gamma - c\overline{\Sigma}}\right].$$

Then, we choose (a, γ) to satisfy

$$\lim_{t\to\infty}\lim_{d\to\infty}\frac{1}{d}\|f_{t+1}(v^{t+1})\|_2^2 = 1, \quad b = 1. \qquad (5.3)$$

The constraint on $\|f_{t+1}(v^{t+1})\|_2^2$ normalizes the GAMP iterate so that, as t grows, its norm does not blow up or vanish. Using state evolution and the characterization of its fixed points, we can show that the conditions (5.3) can be written as

$$\begin{aligned}
1 &= \frac{1}{\mathbb{E}[\overline{\Sigma}]}\mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\overline{\Sigma}]}\overline{G}^2 - 1\right)\frac{\mathcal{T}(\overline{Y})}{a - \mathcal{T}(\overline{Y})}\right]\mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma - \mathbb{E}[\mathcal{T}(\overline{Y})/(a - \mathcal{T}(\overline{Y}))]\overline{\Sigma}}\right], \\
1 &= \frac{1}{\delta}\mathbb{E}\left[\frac{\overline{\Sigma}}{\gamma - \mathbb{E}[\mathcal{T}(\overline{Y})/(a - \mathcal{T}(\overline{Y}))]\overline{\Sigma}}\right].
\end{aligned} \qquad (5.4)$$

Proposition D.4 shows that in the presence of a spectral gap, (5.4) is equivalent to $\zeta(a) = \varphi(a)$, with φ, ζ defined in (3.2) and (3.5). Thus, from (3.6), we have that $(a, \gamma) = (a^*, \gamma(a^*))$.

With the above choice of denoisers, the GAMP iteration can be expressed as

$$\widehat{v}^{t+1} = \frac{D}{a^*\gamma(a^*)}\widehat{v}^t + \widehat{e}^t, \qquad (5.5)$$

for some auxiliary iterate $\widehat{v}^{t+1}$ and error term $\widehat{e}^t$. We show in Section A.5 that $\widehat{e}^t$ asymptotically vanishes as t grows. Now, if $\widehat{e}^t$ is zero, (5.5) is exactly a power iteration for $M := (a^*\gamma(a^*))^{-1}D$. The convergence of this power iteration to the leading eigenvector of M (or, equivalently, of D) crucially relies on the existence of a spectral gap, i.e., on the fact that $\lim_{d\to\infty}\lambda_1(D) > \lim_{d\to\infty}\lambda_2(D)$.

To pinpoint when a spectral gap exists, we establish the limiting value of $\lambda_2(D)$. In Section 5.4, we prove that $\lambda_2(D)$ converges to $\lambda_2 := a^\circ\gamma(a^\circ)$, where a° is given as in (3.4). This is obtained by interlacing the eigenvalues of D with those of a "decoupled" matrix $\widehat{D}$ in which X is replaced with an i.i.d. copy $\widehat{X}$ independent of T. The support of the limiting spectral distribution of $\widehat{D}$ is characterized in [24, Section 3], when T is positive semidefinite. By extending this analysis, we deduce the desired characterization of λ_2. One technical challenge is that, when T is not positive semidefinite, the roles of Σ and T are not interchangeable in determining λ_2, whereas in [24] this symmetry simplifies the arguments.

Given the normalization in (5.3), the largest eigenvalue of M converges to 1 and, thus, $\lim_{d\to\infty}\lambda_1(D) = \lambda_1 := a^*\gamma(a^*)$. Hence, the criticality condition for the existence of a spectral gap reads $a^*\gamma(a^*) > a^\circ\gamma(a^\circ)$. This is equivalent to $a^* > a^\circ$, as adopted in Theorem 3.1, by the monotonicity properties of the function $\psi(a) = a\gamma(a)$ in (3.2) (see Lemma E.1).

To formalize the above reasoning, assume $a^* > a^\circ$ and execute (5.5) for t' steps to amplify the spectral gap:

$$\widehat{v}^{t+t'} \approx M^{t'}\widehat{v}^t, \tag{5.6}$$

where the error terms can be neglected by taking t sufficiently large (and also much larger than t'). Now, we look at the rescaled norms $\|\cdot\|_2/\sqrt{d}$ of both sides of (5.6). As a result of the GAMP state evolution, the rescaled norm of the left-hand side $\|\widehat{v}^{t+t'}\|_2/\sqrt{d}$ can be accurately determined in the high-dimensional limit. Furthermore, it converges to an explicit strictly positive constant in the large t limit, by convergence of state evolution. Thus, inspecting the right-hand side of (5.6) allows us to conclude that $\lambda_1(M)$ must be 1 in the high-dimensional limit. Indeed, if that is not the case, $\|M^{t'}\widehat{v}^t\|_2/\sqrt{d}$ would be either amplified or shrunk geometrically as t' grows, violating the equality in (5.6). At this point, we have

$$\lim_{d\to\infty}\lambda_1(D) = \lambda_1, \quad \lim_{d\to\infty}\lambda_2(D) = \lambda_2,$$

and that $\widehat{v}^t$ is asymptotically aligned with the top eigenvector $v_1(D)$, provided $a^* > a^\circ$. Then, the limiting overlap between β^* and $v_1(D)$ is the same as that between β^* and $\widehat{v}^t$, which is again derived using state evolution.

The rest of this section is organized as follows: Section 5.2 presents the state evolution of GAMP with non-separable denoisers, Section 5.3 establishes its fixed points when GAMP simulates a power iteration, Section 5.4 characterizes the right

edge of the bulk of D, and Section 5.5 puts everything together concluding the proof of Theorem 3.1.

5.2. State evolution of GAMP with non-separable denoisers

In order to precisely state the state evolution result for GAMP, we require the notion of pseudo-Lipschitz functions with matrix-valued inputs and outputs.

Definition 5.1 (Pseudo-Lipschitz functions). A function $h : \mathbb{R}^{k \times m} \to \mathbb{R}^{\ell \times m}$ is called *pseudo-Lipschitz of order j* if there exists L such that

$$\frac{1}{\sqrt{\ell}} \|h(x) - h(y)\|_{\mathrm{F}} \leq \frac{L}{\sqrt{k}} \|x - y\|_{\mathrm{F}}$$
$$\cdot \left[1 + \left(\frac{1}{\sqrt{k}} \|x\|_{\mathrm{F}} \right)^{j-1} + \left(\frac{1}{\sqrt{k}} \|y\|_{\mathrm{F}} \right)^{j-1} \right] \qquad (5.7)$$

for every $x, y \in \mathbb{R}^{k \times m}$.

We will consider sequences of functions $h_i : \mathbb{R}^{k_i \times m} \to \mathbb{R}^{\ell_i \times m}$ indexed by $i \to \infty$ though the index i is often not written explicitly. We call a sequence of functions $(h_i : \mathbb{R}^{k_i \times m} \to \mathbb{R}^{\ell_i \times m})_{i \geq 1}$ *uniformly pseudo-Lipschitz of order j* if there exists a constant L such that for every $i \geq 1$, (5.7) holds. Note that L is a constant as $i \to \infty$.

Define the random vectors

$$\mathfrak{B}^* \sim P^{\otimes d}, \quad \widetilde{\mathfrak{B}}^* = \Sigma^{1/2} \mathfrak{B}^*,$$
$$(G, \varepsilon) \sim \mathcal{N}\left(0_n, \frac{1}{\delta} \mathbb{E}[\bar{\Sigma}] I_n \right) \otimes P_{\varepsilon}^{\otimes n}, \quad Y = q(G, \varepsilon). \qquad (5.8)$$

If $\beta^* \sim \mathrm{Unif}(\sqrt{d}\, \mathbb{S}^{d-1})$, P should be taken to be $\mathcal{N}(0, 1)$.

We further impose the following assumptions which guarantee the existence and finiteness of various state evolution parameters:

(A10) The initializer $\widetilde{v}^0 \in \mathbb{R}^d$ is independent of $\widetilde{X}$. Furthermore,

$$\operatorname*{p-lim}_{d \to \infty} \frac{1}{\sqrt{d}} \|\widetilde{v}^0\|_2$$

exists and is finite. There exists a uniformly pseudo-Lipschitz function $f_0 : \mathbb{R}^d \to \mathbb{R}^d$ of order 1 such that

$$\lim_{d \to \infty} \frac{1}{d} \mathbb{E}\left[\langle f_0(\widetilde{\mathfrak{B}}^*), f_0(\widetilde{\mathfrak{B}}^*) \rangle \right] \leq \operatorname*{p-lim}_{d \to \infty} \frac{1}{d} \|\widetilde{v}^0\|_2^2,$$

and for every uniformly pseudo-Lipschitz $h : \mathbb{R}^d \to \mathbb{R}^d$ of finite order,

$$\operatorname*{p-lim}_{d \to \infty} \frac{1}{d} \langle \widetilde{v}^0, h(\widetilde{\beta}^*) \rangle = \lim_{d \to \infty} \frac{1}{d} \mathbb{E}\left[\langle f_0(\widetilde{\mathfrak{B}}^*), h(\widetilde{\mathfrak{B}}^*) \rangle \right];$$

in particular, limits on both sides of the above two displayed equations exist and are finite. Here, we have set $\widetilde{\beta}^* = \Sigma^{1/2}\beta^*$ and we recall that $\beta^* \sim P^{\otimes d}$ from assumption (A1). Let $\widetilde{\chi} \in \mathbb{R}$, $\widetilde{\sigma}_V \in \mathbb{R}_{\geq 0}$. For any $t \geq 0$,

$$\lim_{d \to \infty} \frac{1}{d} \mathbb{E}\big[\langle f_0(\widetilde{\mathfrak{B}}^*), f_{t+1}(\widetilde{\chi}\widetilde{\mathfrak{B}}^* + \widetilde{\sigma}_V \widetilde{W}_V)\rangle\big]$$

exists and is finite, where $\widetilde{W}_V \sim \mathcal{N}(0_d, I_d)$ is independent of $\widetilde{\mathfrak{B}}^*$.

(A11) Let $\widetilde{\nu} \in \mathbb{R}$, and $T \in \mathbb{R}^{2\times 2}$ be positive definite. For $s, t \geq 0$,

$$\lim_{d \to \infty} \frac{1}{d} \mathbb{E}\big[\langle f_{s+1}(\widetilde{\nu}\widetilde{\mathfrak{B}}^* + \widetilde{N}), f_{t+1}(\widetilde{\nu}\widetilde{\mathfrak{B}}^* + \widetilde{N}')\rangle\big]$$

exists and is finite, where

$$\big(\widetilde{\mathfrak{B}}^*, (\widetilde{N}, \widetilde{N}')\big) \sim \mathcal{N}(0_d, \Sigma) \otimes \mathcal{N}(0_{2d}, T \otimes I_d).$$

Let $\widetilde{\mu} \in \mathbb{R}_{\geq 0}$, and $S \in \mathbb{R}^{2\times 2}$ be positive definite. For any $s, t \geq 0$,

$$\lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[\langle g_s(\widetilde{G} + \widetilde{M}; Y), g_t(\widetilde{G} + \widetilde{M}'; Y)\rangle\big],$$

$$\lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[(\mathrm{div}_g\, g_t(u, q(g, e)))\big|_{u=\widetilde{G}+\widetilde{M}, g=\widetilde{G}, e=\varepsilon}\big]$$

exist and are finite, where

$$(\widetilde{G}, \varepsilon, \widetilde{M}, \widetilde{M}') \sim \mathcal{N}(0_n, \widetilde{\mu}^2 I_n) \otimes P_\varepsilon^{\otimes n} \otimes \mathcal{N}(0_{2n}, S \otimes I_n) \quad \text{and} \quad Y = q(\widetilde{G}, \varepsilon).$$

The state evolution result – formally stated below – asserts that, for each $t \geq 0$, in the large n limit, the joint distributions of the AMP iterates $(\widetilde{\beta}^*, v^1, v^2, \ldots, v^{t+1})$ and $(g = X\beta^*, u^0, u^1, \ldots, u^t)$ converge to the laws of $(\widetilde{\mathfrak{B}}^*, V_1, V_2, \ldots, V_{t+1})$ and $(G, U_0, U_1, \ldots, U_t)$, respectively. For $t \geq 0$, the random vectors $U_t \in \mathbb{R}^n$ and $V_{t+1} \in \mathbb{R}^d$ are defined as:

$$U_t = \mu_t G + \sigma_{U,t} W_{U,t}, \quad V_{t+1} = \chi_{t+1}\widetilde{\mathfrak{B}}^* + \sigma_{V,t+1} W_{V,t+1}, \tag{5.9}$$

where $W_{U,t} \sim \mathcal{N}(0_n, I_n)$ is independent of (G, ε), and $W_{V,t+1} \sim \mathcal{N}(0_d, I_d)$ is independent of $\widetilde{\mathfrak{B}}^*$. The constants $\mu_t, \sigma_{U,t}, \chi_{t+1}, \sigma_{V,t+1}$ are recursively defined, starting from

$$\mu_0 = \frac{\delta}{\mathbb{E}[\overline{\Sigma}]} \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[\langle \widetilde{\mathfrak{B}}^*, f_0(\widetilde{\mathfrak{B}}^*)\rangle\big],$$

$$\sigma_{U,0}^2 = \operatorname*{p\text{-}lim}_{n \to \infty} \frac{1}{n} \langle \widetilde{v}^0, \widetilde{v}^0\rangle - \frac{\mathbb{E}[\overline{\Sigma}]}{\delta} \mu_0^2. \tag{5.10}$$

For $t \geq 0$, we have

$$
\chi_{t+1} = \frac{\delta}{\mathbb{E}[\overline{\Sigma}]} \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\big[\langle G, g_t(U_t; Y)\rangle\big] - \mu_t \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\big[\mathrm{div}_{U_t}\, g_t(U_t; Y)\big],
$$

$$
\sigma_{V,t+1}^2 = \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\big[\langle g_t(U_t; Y), g_t(U_t; Y)\rangle\big],
$$

$$(5.11)$$

and

$$
\mu_{t+1} = \frac{\delta}{\mathbb{E}[\overline{\Sigma}]} \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\big[\langle \widetilde{\mathfrak{B}}^*, f_t(V_{t+1})\rangle\big],
$$

$$
\sigma_{U,t+1}^2 = \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\big[\langle f_t(V_{t+1}), f_t(V_{t+1})\rangle\big] - \frac{\mathbb{E}[\overline{\Sigma}]}{\delta} \mu_{t+1}^2.
$$

$$(5.12)$$

The two sequences of random vectors $(W_{U,t})_{t\geq 0}$ and $(W_{V,t+1})_{t\geq 0}$, are each jointly Gaussian with the following laws:

$$
\begin{bmatrix} \sigma_{U,0} W_{U,0} \\ \sigma_{U,1} W_{U,1} \\ \vdots \\ \sigma_{U,t} W_{U,t} \end{bmatrix} \sim \mathcal{N}\big(0_{(t+1)n}, \Phi_t \otimes I_n\big), \qquad \begin{bmatrix} \sigma_{V,1} W_{V,1} \\ \sigma_{V,2} W_{V,2} \\ \vdots \\ \sigma_{V,t+1} W_{V,t+1} \end{bmatrix} \sim \mathcal{N}\big(0_{(t+1)d}, \Psi_t \otimes I_d\big),
$$

$$(5.13)$$

where $\Phi_t, \Psi_t \in \mathbb{R}^{(t+1)\times(t+1)}$ are matrices with entries:

$$
(\Phi_t)_{1,1} := \underset{n\to\infty}{\mathrm{p\text{-}lim}} \frac{1}{n} \langle \tilde{v}^0, \tilde{v}^0 \rangle - \frac{\mathbb{E}[\overline{\Sigma}]}{\delta} \mu_0^2,
$$

$$
(\Phi_t)_{1,s+1} := \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\big[\langle f_0(\widetilde{\mathfrak{B}}^*) - \mu_0 \widetilde{\mathfrak{B}}^*, f_s(V_s) - \mu_s \widetilde{\mathfrak{B}}^* \rangle\big] \quad \text{for } 1 \leq s \leq t,
$$

$$(5.14)$$

$$
(\Phi_t)_{r+1,s+1} := \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\big[\langle f_r(V_r) - \mu_r \widetilde{\mathfrak{B}}^*, f_s(V_s) - \mu_s \widetilde{\mathfrak{B}}^* \rangle\big] \quad \text{for } 1 \leq r, s \leq t,
$$

$$(5.15)$$

$$
(\Psi_t)_{r+1,s+1} := \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\big[\langle g_r(U_r; Y), g_s(U_s; Y)\rangle\big] \qquad \text{for } 0 \leq r, s \leq t.
$$

$$(5.16)$$

Note that for $r = s$, $(\Psi_t)_{r+1,r+1} = \sigma_{V,r+1}^2$ is consistent with (5.11) and

$$
\begin{aligned}
(\Phi_t)_{r+1,r+1} &= \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\big[\langle f_r(V_r) - \mu_r \widetilde{\mathfrak{B}}^*, f_r(V_r) - \mu_r \widetilde{\mathfrak{B}}^* \rangle\big] \\
&= \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\big[\langle f_r(V_r), f_r(V_r)\rangle\big] \\
&\quad - 2\mu_r \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\big[\langle f_r(V_r), \widetilde{\mathfrak{B}}^* \rangle\big] + \mu_r^2 \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\big[\langle \widetilde{\mathfrak{B}}^*, \widetilde{\mathfrak{B}}^* \rangle\big] \\
&= \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\big[\langle f_r(V_r), f_r(V_r)\rangle\big] - 2\mu_r^2 \frac{\mathbb{E}[\overline{\Sigma}]}{\delta} + \mu_r^2 \frac{\mathbb{E}[\overline{\Sigma}]}{\delta} = \sigma_{U,r}^2
\end{aligned}
$$

is consistent with (5.12), where the last line above follows from the definition of μ_r in (5.12). As $(G, W_{U,0}, \ldots, W_{U,t})$ are jointly Gaussian by (5.13), their covariance structure (and therefore that of $(G, U_0, \ldots, U_t)$ in view of (5.9)) is completely determined by the constants defined in (5.10), (5.11), (5.12), (5.14), and (5.15). Similarly $(V_1 - \chi_1 \widetilde{\mathfrak{B}}^*, \ldots, V_{t+1} - \chi_{t+1} \widetilde{\mathfrak{B}}^*) = (\sigma_{V,1} W_{V,1}, \ldots, \sigma_{V,t+1} W_{V,t+1})$ are jointly Gaussian by (5.9) and (5.13), hence the covariance structure of $(\widetilde{\mathfrak{B}}^*, V_1, \ldots, V_{t+1})$ is completely determined by the constants in (5.11) and (5.16).

We are now ready to present the state evolution result. Its proof, deferred to Section A.1, reduces the GAMP iteration in (5.1) to a family of abstract AMP algorithms introduced in [42] for which a general state evolution result has been established. In the abstract AMP algorithm, iterates are associated with the edges of a given directed graph, and the denoising functions are allowed to be non-separable, as needed in our case.

Proposition 5.1 (State evolution). *Consider the GLM in Section 2.1 subject to assumptions (A1) to (A4) and the GAMP iteration in (5.1). Let initializers $\widetilde{u}^{-1} = 0_n$ and $\widetilde{v}^0 \in \mathbb{R}^d$ satisfy assumption (A10). For every $t \geq 0$, let $(g_t: \mathbb{R}^{2n} \to \mathbb{R}^n)_{n \geq 1}$ and $(f_{t+1}: \mathbb{R}^d \to \mathbb{R}^d)_{d \geq 1}$ be uniformly pseudo-Lipschitz functions of finite constant order subject to assumption (A11). For any $t \geq 0$, let $(h_1: \mathbb{R}^{n(t+2)} \to \mathbb{R})_{n \geq 1}$ and $(h_2: \mathbb{R}^{d(t+2)} \to \mathbb{R})_{d \geq 1}$ be two sequences of uniformly pseudo-Lipschitz test functions of finite order. Then,*

$$\underset{n \to \infty}{\text{p-lim}}\, h_1(g, u^0, u^1, \ldots, u^t) - \mathbb{E}[h_1(G, U_0, U_1, \ldots, U_t)] = 0,$$
$$\underset{d \to \infty}{\text{p-lim}}\, h_2(\widetilde{\beta}^*, v^1, v^2, \ldots, v^{t+1}) - \mathbb{E}[h_2(\widetilde{\mathfrak{B}}^*, V_1, V_2, \ldots, V_{t+1})] = 0,$$

(5.17)

where $(U_t, V_{t+1})_{t \geq 0}$ are given in (5.9).

5.3. GAMP as a power method and its fixed points

We now formalize the argument in Section 5.1. Recall the definition of $\overline{Y}$ in (2.2) and $s(\cdot)$ in (3.1). Let

$$\mathcal{A} := \{(a, \gamma) : a > \sup \text{supp}(\mathcal{T}(\overline{Y})), \gamma > s(a)\}$$

and $(a^*, \gamma(a^*)) \in \mathcal{A}$ be defined through (3.6), where the largest solution a^* is taken. For convenience, for the rest of the paper, we will use the shorthand

$$\gamma^* := \gamma(a^*), \quad \gamma^\circ := \gamma(a^\circ).$$

(5.18)

If $a^* > a°$ (where $a°$ is defined in (3.4)), Proposition D.4 shows that this pair of equations is equivalent to (5.4). Furthermore, let

$$\mathcal{F}_a(\cdot) = \frac{\mathcal{T}(\cdot)}{a - \mathcal{T}(\cdot)}, \quad a > \sup \text{supp}(\mathcal{T}(\bar{Y})),$$

$$F = \text{diag}(\mathcal{F}_{a^*}(y)), \quad c = \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]. \tag{5.19}$$

Let us initialize the iteration in (5.1) with $\tilde{u}^{-1} = 0_n$ and $\tilde{v}^0 \in \mathbb{R}^d$ defined in (5.30), and for subsequent iterates, set

$$g_t(u^t; y) = Fu^t, \quad f_{t+1}(v^{t+1}) = (\gamma_{t+1}I_d - c\Sigma)^{-1}\Sigma v^{t+1}, \quad t \geq 0. \tag{5.20}$$

Recall from assumption (A5) that $\mathcal{T}: \mathbb{R} \to \mathbb{R}$ is bounded and pseudo-Lipschitz of finite order. Since $a^* > \sup \text{supp}(\mathcal{T}(\bar{Y}))$, $\mathcal{F}_{a^*}: \mathbb{R} \to \mathbb{R}$ is also bounded and pseudo-Lipschitz of finite order. Therefore, for every $t \geq 0$, $(g_t: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n)_{n \geq 1}$ is a sequence of uniformly pseudo-Lipschitz functions of finite order in both arguments. The parameter $\gamma_{t+1} \in (s(a^*), \infty)$ is such that

$$\underset{d \to \infty}{\text{p-lim}} \frac{1}{d} \|f_{t+1}(v^{t+1})\|_2^2 = \lim_{d \to \infty} \frac{1}{d} \mathbb{E}\left[\|f_{t+1}(V_{t+1})\|_2^2\right] = 1 \tag{5.21}$$

for $t \geq 0$. The first equality above follows from the state evolution result in Proposition 5.1. For notational convenience, let

$$B_{t+1} := (\gamma_{t+1}I_d - c\Sigma)^{-1}\Sigma. \tag{5.22}$$

Since $\gamma_{t+1} > s(a^*)$ and $\|\Sigma\|_2$ is uniformly bounded by assumption (A2), $\|B_{t+1}\|_2$ is uniformly bounded. Therefore for every $t \geq 0$, $(f_{t+1}: \mathbb{R}^d \to \mathbb{R}^d)_{d \geq 1}$ is a sequence of pseudo-Lipschitz functions of order 1.

With the above definitions, the Onsager coefficients become

$$c_t = \frac{1}{n}\text{Tr}(F), \quad b_{t+1} = \frac{d}{n}\text{Tr}(B_{t+1}), \tag{5.23}$$

for every $t \geq 0$. Furthermore, the state evolution in (5.12) and (5.11) specializes to the following recursion

$$\mu_t = \frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\lim_{n \to \infty}\frac{1}{n}\mathbb{E}\left[(\tilde{\mathfrak{B}}^*)^\top B_t V_t\right],$$

$$\sigma_{U,t}^2 = \lim_{n \to \infty}\frac{1}{n}\mathbb{E}\left[V_t^\top B_t^\top B_t V_t\right] - \frac{\mathbb{E}[\bar{\Sigma}]}{\delta}\mu_t^2,$$

$$\chi_{t+1} = \frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\lim_{n \to \infty}\frac{1}{n}\mathbb{E}\left[G^\top \text{diag}(\mathcal{F}_{a^*}(Y))U_t\right] - \mu_t\mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})],$$

$$\sigma_{V,t+1}^2 = \lim_{n \to \infty}\frac{1}{n}\mathbb{E}\left[U_t^\top \text{diag}(\mathcal{F}_{a^*}(Y))^2 U_t\right]. \tag{5.24}$$

Let

$$z_1 := \mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2}\right], \quad z_2 := \mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2}\right]. \quad (5.25)$$

Note that $z_1, z_2 > 0$. Recalling w_1, w_2 from (3.9), and (3.10), define

$$\chi = \sqrt{\frac{1 - w_2}{(1 - w_2)z_1 + w_1 z_2}}, \quad \sigma_V = \sqrt{\frac{w_1}{(1 - w_2)z_1 + w_1 z_2}}, \quad (5.26)$$

$$\mu = \frac{1}{\mathbb{E}[\bar{\Sigma}]}\mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma}}\right]\sqrt{\frac{1 - w_2}{(1 - w_2)z_1 + w_1 z_2}}, \quad (5.27)$$

$$\sigma_U = \sqrt{\frac{1/\delta}{(1 - w_2)z_1 + w_1 z_2}}$$
$$\cdot\left(\mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2}\right] - \frac{1}{\mathbb{E}[\bar{\Sigma}]}\mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma}}\right]^2\right.$$
$$\left. + \frac{1}{\mathbb{E}[\bar{\Sigma}]^2}\mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2}\right]\mathbb{E}[\bar{G}^2\mathcal{F}_{a^*}(\bar{Y})^2]\mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma}}\right]^2\right)^{1/2}. \quad (5.28)$$

Note that all these quantities are well defined provided $a^* > a^\circ$. Indeed, $w_1 > 0$ and $1 - w_2 > 0$ under the latter condition. Also, the second factor in the definition of σ_U is positive since the sum of the first two terms is non-negative by Cauchy–Schwarz and the third term is positive. Define also $\gamma^\sharp$ as the unique solution in $(s(a^*), \infty)$ to

$$1 = \frac{1}{\delta}\mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})^2]\mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma^\sharp - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2}\right]. \quad (5.29)$$

The well-posedness of $\gamma^\sharp$ follows the same reasoning after (3.3).

We now characterize the fixed points of state evolution and show that the recursion can be initialized precisely at the fixed point. The proof of the next two lemmas are obtained via a series of manipulations which are deferred to Sections A.2 and A.3.

Lemma 5.2 (Fixed points of state evolution). *The quintuple*

$$(\mu_t, \sigma_{U,t}, \chi_{t+1}, \sigma_{V,t+1}, \gamma_{t+1})$$

in the recursion given by (5.24) *and* (5.21) *has 3 fixed points* $\mathsf{FP}_+, \mathsf{FP}_-, \mathsf{FP}_0 \in \mathbb{R}^5$:

$$\mathsf{FP}_+ = (\mu, \sigma_U, \chi, \sigma_V, \gamma^*), \quad \mathsf{FP}_- = (-\mu, \sigma_U, -\chi, \sigma_V, \gamma^*),$$
$$\mathsf{FP}_0 = \left(0, \frac{1}{\sqrt{\delta}}, 0, \mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma^\sharp - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2}\right]^{-1/2}, \gamma^\sharp\right),$$

where the parameters on the right are given in (5.18) *and* (5.26) *to* (5.29).

We initialize the AMP iteration with

$$\widetilde{u}^{-1} = 0_n, \quad \widetilde{v}^0 := \mu\,\widetilde{\beta}^* + \sqrt{1 - \mu^2 \mathbb{E}[\bar{\Sigma}]}\, w \in \mathbb{R}^d, \tag{5.30}$$

where we have set $\widetilde{\beta}^* = \Sigma^{1/2}\beta^*$, $w \sim \mathcal{N}(0_d, I_d)$ is independent of everything else and μ is given in (5.27). This choice is valid since from the proof of Lemma 5.3 one can deduce that $1 - \mu^2 \mathbb{E}[\bar{\Sigma}] > 0$. The scaling ensures that p-$\lim_{d\to\infty} \|\widetilde{v}^0\|_2^2/d = 1$ almost surely. According to (5.10), (5.30) gives that the state evolution parameters are initialized as

$$\mu_0 = \frac{\delta}{\mathbb{E}[\bar{\Sigma}]} \lim_{n\to\infty} \frac{\mu}{n}\mathbb{E}\big[\langle \widetilde{\mathfrak{B}}^*, \widetilde{\mathfrak{B}}^* \rangle\big] = \mu,$$

$$\sigma_{U,0}^2 = \underset{n\to\infty}{\text{p-lim}} \frac{1}{n}\langle \widetilde{v}^0, \widetilde{v}^0 \rangle - \frac{\mathbb{E}[\bar{\Sigma}]}{\delta}\mu_0^2 = \frac{1}{\delta} - \frac{\mathbb{E}[\bar{\Sigma}]}{\delta}\mu^2. \tag{5.31}$$

Lemma 5.3 (State evolution stays put). *Initialized with* (5.31) *above, the parameters* $(\mu_t, \sigma_{U,t}, \chi_{t+1}, \sigma_{V,t+1})_{t\geq 0}$ *of the state evolution recursion in* (5.24) *and* (5.21) *stay at the initialization, that is, for every* $t \geq 0$:

$$\mu_t = \mu, \quad \sigma_{U,t} = \sigma_U, \quad \chi_{t+1} = \chi, \quad \sigma_{V,t+1} = \sigma_V, \quad \gamma_{t+1} = \gamma^*,$$

where the right-hand sides are defined in (5.18) *and* (5.26) *to* (5.28).

5.4. Right edge of the bulk of D

Let

$$\widehat{D} = \Sigma^{1/2}\widehat{X}^\top T \widehat{X}\Sigma^{1/2} \text{ with } T = \text{diag}\big(\mathcal{T}(y)\big) = \text{diag}\big(\mathcal{T}(q(\widetilde{X}\Sigma^{1/2}\beta^*, \varepsilon))\big), \tag{5.32}$$

and $\widehat{X} \in \mathbb{R}^{n\times d}$ has i.i.d. $\mathcal{N}(0, 1/n)$ entries, independent of T. One should think of $\widehat{D}$ as a "decoupled" version of D in the sense that $\widehat{X}$ and T are independent and no outlier eigenvalue is expected to show up in the spectrum of $\widehat{D}$. This is to be contrasted with $D = \Sigma^{1/2}\widetilde{X}^\top T \widetilde{X}\Sigma^{1/2}$ in which T depends on $\widetilde{X}$ (see (5.32)), and the top eigenvalue of D will be detached from the bulk of the spectrum provided that $a^* > a^\circ$.

Given the above intuition, one expects that the behavior of the right edge of the bulk of $\widehat{D}$ resembles that of D. This is made formal in the following lemma, which is proved in Section A.4. The idea is to first show that $\lambda_3(\widehat{D}) \leq \lambda_2(D) \leq \lambda_1(\widehat{D})$ using the variational representation of eigenvalues, and then use [38, 98] to show that both $\lambda_1(\widehat{D})$ and $\lambda_3(\widehat{D})$ converge to the right edge of the bulk of $\widehat{D}$. We comment on the second step. Building on the almost sure weak convergence result of the empirical spectral distribution of $\widehat{D}$ [98, Theorem 1.2.1], it was proved in [73, Theorem 1] that almost surely there exists no eigenvalue outside the support of the limiting spectral

distribution, and [24, Section 3] further characterized the support. However, both [73] and [24] assumed a positive semidefinite T which corresponds to $\mathcal{T} \geq 0$. Thus, we build on [98, Theorem 1.2.1] and use a recent strong asymptotic freeness result of GOE and deterministic matrices [38, Theorem 4.3] which guarantees the absence of eigenvalues outside the support of the limiting spectral distribution. Of particular benefit to our purposes is that neither [98, Theorem 1.2.1] nor [38, Theorem 4.3] requires T to be PSD.

Lemma 5.4. *Consider the matrices D and $\widehat{D}$ in (1.2) and (5.32), respectively. Denote by $\bar{\mu}_{\widehat{D}}$ the limiting spectral distribution of $\widehat{D}$. Then, we have*

$$\lim_{d \to \infty} \lambda_2(D) = \sup \operatorname{supp}(\bar{\mu}_{\widehat{D}}) \quad \text{almost surely.}$$

Next, we characterize the right edge of the support of $\bar{\mu}_{\widehat{D}}$. The detailed proof of the lemma below is given in Section E, and it generalizes the analysis in [24, Section 3], showing that the same characterization of the support therein also holds for a possibly non-positive T (or equivalently $\mathcal{T}$). The critical obstacle for non-positive T is that the Stieltjes-like transform $z \mapsto \mathbb{E}[\mathcal{T}(\bar{Y})/(\mathcal{T}(\bar{Y}) - z)]$ no longer maps the complex upper-half plane into itself, rendering parts of [24] using this property unusable. We treat this problem by considering meromorphic generalizations of various concepts in [24] (e.g., Proposition E.8 in Section E plays the role of [24, Proposition 1.2]).

Lemma 5.5. *Let $a^\circ > \sup \operatorname{supp}(\mathcal{T}(\bar{Y}))$ be the largest critical point of ψ. Then, we have*

$$\sup \operatorname{supp}(\bar{\mu}_{\widehat{D}}) = \psi(a^\circ). \tag{5.33}$$

5.5. Concluding the proof of Theorem 3.1

In this final section, we show the following lemma.

Lemma 5.6. *Consider v^t obtained from the GAMP iteration in (5.1) with denoisers in (5.19) and (5.20) and initializers in (5.30). Let $\widehat{v}^t := \Sigma^{-1/2}(\gamma^* I_d - c\Sigma)^{-1}\Sigma v^t$. If $a^* > a^\circ$, then*

$$\lim_{t \to \infty} \operatorname{p-lim}_{d \to \infty} \frac{\langle \widehat{v}^t, v_1(D) \rangle^2}{\|\widehat{v}^t\|_2^2} = 1, \quad \operatorname{p-lim}_{d \to \infty} \lambda_1(D) = \lambda_1 > \lambda_2, \tag{5.34}$$

where λ_1, λ_2 are defined in (3.7).

Then, (5.34) directly gives the first part of (3.11) in Theorem 3.1; the second part follows from Lemmas 5.4 and 5.5; and the expression in (3.12) for the overlap is a consequence of state evolution, whose proof is given at the end of this section.

Proof of Lemma 5.6. Recall the following definitions: B_{t+1} in (5.22), $f_{t+1}(v^{t+1}) = B_{t+1}v^{t+1}$ (see (5.20)) and $c = \mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})]$ (see (5.19)). Let

$$b := \frac{1}{\delta}\mathbb{E}\left[\frac{\bar{\Sigma}}{\gamma^* - c\bar{\Sigma}}\right], \quad B := (\gamma^* I_d - c\Sigma)^{-1}\Sigma, \tag{5.35}$$

be the fixed points of b_{t+1}, B_{t+1}, respectively, where γ^* (together with a^*) satisfies (5.4). Note that $b = 1$ by (3.3). For $t \geq 1$, define

$$\begin{aligned} e_1^t &:= u^t - u^{t-1} \in \mathbb{R}^n, \\ e_2^t &:= v^{t+1} - v^t \in \mathbb{R}^d. \end{aligned} \tag{5.36}$$

The GAMP iteration in (5.1) can be written as

$$\begin{aligned} u^t &= \tilde{X}B_t v^t - b_t F u^{t-1}, \\ v^{t+1} &= \tilde{X}^\top F u^t - c_t B_t v^t. \end{aligned}$$

Using the first equation in the second, we get

$$v^{t+1} = (\tilde{X}^\top F\tilde{X} - c_t I_d)B_t v^t - b_t \tilde{X}^\top F^2 u^{t-1}. \tag{5.37}$$

Using the definition of e_1^t in the iteration for u^t, we have

$$u^{t-1} = \tilde{X}B_t v^t - b_t F u^{t-1} - e_1^t.$$

Solving for u^{t-1} yields

$$u^{t-1} = (b_t F + I_n)^{-1}\tilde{X}B_t v^t - (b_t F + I_n)^{-1}e_1^t.$$

Then, we can eliminate u^{t-1} in the iteration for v^{t+1} by substituting the right-hand side above in (5.37) and, after some manipulations, we obtain

$$v^{t+1} = \left[\tilde{X}^\top F(b_t F + I_n)^{-1}\tilde{X} - c_t I_d\right]B_t v^t + b_t \tilde{X}^\top F^2(b_t F + I_n)^{-1}e_1^t.$$

We expand b_t and B_t respectively around their fixed points b and B to write

$$\begin{aligned} v^{t+1} = &\left[\tilde{X}^\top F(bF + I_n)^{-1}\tilde{X} - cI_d\right]Bv^t \\ &+ (b - b_t)\tilde{X}^\top F(b_t F + I_n)^{-1}(bF + I_n)^{-1}\tilde{X}B_t v^t \\ &+ (\gamma^* - \gamma_t)\tilde{X}^\top F(bF + I_n)^{-1}\tilde{X}(\gamma_t I_d - c\Sigma)^{-1}(\gamma^* I_d - c\Sigma)^{-1}\Sigma v^t \\ &+ (c_t - c)B_t v^t + c(\gamma_t - \gamma^*)(\gamma_t I_d - c\Sigma)^{-1}(\gamma^* I_d - c\Sigma)^{-1}\Sigma v^t \\ &+ b_t \tilde{X}^\top F^2(b_t F + I_n)^{-1}e_1^t. \end{aligned}$$

Using the definition of e_2^t, we further have

$$
\begin{aligned}
(I_d + cB)v^{t+1} &= \tilde{X}^\top F(bF + I_n)^{-1}\tilde{X}Bv^t + cBe_2^t \\
&\quad + (b - b_t)\tilde{X}^\top F(b_t F + I_n)^{-1}(bF + I_n)^{-1}\tilde{X}B_t v^t \\
&\quad + (\gamma^* - \gamma_t)\tilde{X}^\top F(bF + I_n)^{-1}\tilde{X}(\gamma_t I_d - c\Sigma)^{-1}(\gamma^* I_d - c\Sigma)^{-1}\Sigma v^t \\
&\quad + (c_t - c)B_t v^t + c(\gamma_t - \gamma^*)(\gamma_t I_d - c\Sigma)^{-1}(\gamma^* I_d - c\Sigma)^{-1}\Sigma v^t \\
&\quad + b_t \tilde{X}^\top F^2(b_t F + I_n)^{-1}e_1^t.
\end{aligned} \tag{5.38}
$$

Define $e^t \in \mathbb{R}^d$ by

$$
\begin{aligned}
e^t &:= c\Sigma^{1/2}Be_2^t + (b - b_t)\Sigma^{1/2}\tilde{X}^\top F(b_t F + I_n)^{-1}(bF + I_n)^{-1}\tilde{X}B_t v^t \\
&\quad + (\gamma^* - \gamma_t)\Sigma^{1/2}\tilde{X}^\top F(bF + I_n)^{-1}\tilde{X}(\gamma_t I_d - c\Sigma)^{-1}(\gamma^* I_d - c\Sigma)^{-1}\Sigma v^t \\
&\quad + (c_t - c)\Sigma^{1/2}B_t v^t + c(\gamma_t - \gamma^*)\Sigma^{1/2}(\gamma_t I_d - c\Sigma)^{-1}(\gamma^* I_d - c\Sigma)^{-1}\Sigma v^t \\
&\quad + b_t \Sigma^{1/2}\tilde{X}^\top F^2(b_t F + I_n)^{-1}e_1^t.
\end{aligned} \tag{5.39}
$$

Multiplying both sides of (5.38) by $\Sigma^{1/2}$, we arrive at

$$
\Sigma^{1/2}(I_d + cB)v^{t+1} = \Sigma^{1/2}\tilde{X}^\top F(bF + I_n)^{-1}\tilde{X}Bv^t + e^t.
$$

By the definition of D (see (1.2)) and the choice of $\mathcal{F}_{a^*}$ (see (5.19)), we note that

$$
\Sigma^{1/2}\tilde{X}^\top F(bF + I_n)^{-1}\tilde{X}\Sigma^{1/2} = \frac{1}{a^*}D
$$

(recall from (5.35) that $b = 1$). Also, by the definition of B (see (5.35)), we have the identity

$$
\frac{1}{\gamma^*}\Sigma^{1/2}(I_d + cB) = \Sigma^{-1/2}B, \tag{5.40}
$$

both sides of which we define to be $\tilde{B} \in \mathbb{R}^{d \times d}$. Using the above observations and letting

$$
\hat{v}^{t+1} := \tilde{B}v^{t+1} \in \mathbb{R}^d, \tag{5.41}
$$

we obtain

$$
\hat{v}^{t+1} = M\hat{v}^t + \frac{1}{\gamma^*}e^t, \quad \text{where } M := \frac{D}{\lambda_1}, \lambda_1 := a^*\gamma^*, \tag{5.42}
$$

which takes the form of a power iteration with an error term.

It is now convenient to shift the spectrum of M to the right so that all of its eigenvalues are positive. Specifically, choose $\ell > 0$ to be a sufficiently large constant. By (A.66), it suffices to take $\ell = C_D + 1 > \|D\|_2 + 1$, where the constant $C_D \in (0, \infty)$

is defined in (A.65). Adding $\frac{\ell}{\lambda_1}\hat{v}^{t+1}$ on both sides of (5.42) and using the definitions of $\hat{v}^t$ in (5.41) and e_2^t in (5.36), we have

$$\left(1 + \frac{\ell}{\lambda_1}\right)\hat{v}^{t+1} = \frac{D + \ell I_d}{\lambda_1}\hat{v}^t + \frac{\ell}{\lambda_1}\tilde{B}e_2^t + \frac{1}{\gamma^*}e^t.$$

Using the following notation:

$$\hat{M} := \frac{D + \ell I_d}{\lambda_1 + \ell}, \quad \hat{e}^t := \frac{\ell}{\lambda_1 + \ell}\tilde{B}e_2^t + \frac{a^*}{\lambda_1 + \ell}e^t, \tag{5.43}$$

we write the iteration as

$$\hat{v}^{t+1} = \hat{M}\hat{v}^t + \hat{e}^t. \tag{5.44}$$

By construction, $\hat{M}$ is strictly positive definite, and all results concerning the spectral properties of $\hat{M}$ can be easily translated to those of M by canceling the shift ℓ.

Suppose that the iteration in (5.44) has been run for a certain large constant $t > 0$ steps. We further run it for an additional t' steps for some large constant $t' > 0$. By unrolling the iteration down to time t, we obtain

$$\hat{v}^{t+t'} = \hat{M}^{t'}\hat{v}^t + \hat{e}^{t,t'}, \tag{5.45}$$

where

$$\hat{e}^{t,t'} := \sum_{s=1}^{t'} \hat{M}^{t'-s}\hat{e}^{t+s-1}. \tag{5.46}$$

Taking the normalized squared norm $\frac{1}{d}\|\cdot\|_2^2$ on both sides of (5.45) and sending first d, then t, and finally t' to infinity, we get the left-hand side

$$\lim_{t'\to\infty}\lim_{t\to\infty}\text{p-lim}_{d\to\infty}\frac{1}{d}\|\hat{v}^{t+t'}\|_2^2 = \lim_{t'\to\infty}\lim_{t\to\infty}\text{p-lim}_{d\to\infty}\frac{1}{d}\|\tilde{B}v^{t+t'}\|_2^2$$

$$= \lim_{t'\to\infty}\lim_{t\to\infty}\text{p-lim}_{d\to\infty}\frac{1}{d}\|\Sigma^{-1/2}(\gamma^* I_d - c\Sigma)^{-1}\Sigma v^{t+t'}\|_2^2$$

$$= \lim_{t'\to\infty}\lim_{t\to\infty}\lim_{d\to\infty}\frac{1}{d}\mathbb{E}\big[\|\Sigma^{-1/2}(\gamma^* I_d - c\Sigma)^{-1}\Sigma V_{t+t'}\|_2^2\big]$$

$$= \lim_{t'\to\infty}\lim_{t\to\infty}\lim_{d\to\infty}\frac{1}{d}\mathbb{E}\big[\|\Sigma^{-1/2}(\gamma^* I_d - c\Sigma)^{-1}\Sigma\tilde{\mathfrak{B}}^*\|_2^2\big]\chi_{t+t'}^2$$

$$+ \frac{1}{d}\mathbb{E}\big[\|\Sigma^{-1/2}(\gamma^* I_d - c\Sigma)^{-1}\Sigma W_{V,t+t'}\|_2^2\big]\sigma_{V,t+t'}^2$$

$$= \lim_{t'\to\infty}\lim_{t\to\infty}\lim_{d\to\infty}$$

$$\Big(\frac{1}{d}\mathbb{E}\big[\mathfrak{B}^{*\top}\Sigma^{1/2}\Sigma(\gamma^* I_d - c\Sigma)^{-1}\Sigma^{-1}(\gamma^* I_d - c\Sigma)^{-1}\Sigma\Sigma^{1/2}\mathfrak{B}^*\big]\chi_{t+t'}^2$$

$$+ \frac{1}{d}\mathbb{E}\big[W_{V,t+t'}^\top\Sigma(\gamma^* I_d - c\Sigma)^{-1}\Sigma^{-1}(\gamma^* I_d - c\Sigma)^{-1}\Sigma W_{V,t+t'}\big]\sigma_{V,t+t'}^2\Big)$$

$$\begin{aligned}
&= \lim_{t' \to \infty} \lim_{t \to \infty} \mathbb{E}\left[\frac{\overline{\Sigma}^2}{(\gamma^* - c\overline{\Sigma})^2}\right]\chi^2_{t+t'} + \mathbb{E}\left[\frac{\overline{\Sigma}}{(\gamma^* - c\overline{\Sigma})^2}\right]\sigma^2_{V,t+t'} \\
&= \mathbb{E}\left[\frac{\overline{\Sigma}^2}{(\gamma^* - c\overline{\Sigma})^2}\right]\chi^2 + \mathbb{E}\left[\frac{\overline{\Sigma}}{(\gamma^* - c\overline{\Sigma})^2}\right]\sigma^2_V =: \nu^2,
\end{aligned} \tag{5.47}$$

where we use the state evolution result (Proposition 5.1) in the third equality. Taking $\frac{1}{d}\|\cdot\|_2^2$ and the same sequential limits on the right-hand side, we have

$$\lim_{t' \to \infty} \lim_{t \to \infty} \text{p-lim}_{d \to \infty} \frac{1}{d}\|\widehat{M}^{t'}\widehat{v}^t + \widehat{e}^{t,t'}\|_2^2. \tag{5.48}$$

We claim that

$$\lim_{t' \to \infty} \lim_{t \to \infty} \text{p-lim}_{d \to \infty} \frac{1}{d}\|\widehat{e}^{t,t'}\|_2^2 = 0, \tag{5.49}$$

which implies, by the triangle inequality, that (5.48) is equal to

$$\lim_{t' \to \infty} \lim_{t \to \infty} \text{p-lim}_{d \to \infty} \frac{1}{d}\|\widehat{M}^{t'}\widehat{v}^t\|_2^2. \tag{5.50}$$

The proof of (5.49) requires the technical analysis of various error terms, and it is deferred to Section A.5. The quantity in (5.50) can be decomposed as

$$\begin{aligned}
\frac{1}{d}\|\widehat{M}^{t'}\widehat{v}^t\|_2^2 &= \frac{1}{d}\|\widehat{M}^{t'}(\Pi + \Pi^\perp)\widehat{v}^t\|_2^2 \\
&= \frac{1}{d}\|\widehat{M}^{t'}\Pi\widehat{v}^t\|_2^2 + \frac{1}{d}\|\widehat{M}^{t'}\Pi^\perp\widehat{v}^t\|_2^2 + \frac{2}{d}\langle\widehat{M}^{t'}\Pi\widehat{v}^t, \widehat{M}^{t'}\Pi^\perp\widehat{v}^t\rangle,
\end{aligned} \tag{5.51}$$

where $\Pi := v_1(D)v_1(D)^\top$ and $\Pi^\perp := I_d - \Pi$. Note that the eigendecomposition of $\widehat{M}^{t'}$ is

$$\widehat{M}^{t'} = \sum_{i=1}^d \lambda_i(\widehat{M}^{t'})v_i(\widehat{M}^{t'})v_i(\widehat{M}^{t'})^\top = \sum_{i=1}^d \lambda_i(\widehat{M})^{t'}v_i(D)v_i(D)^\top,$$

since for any univariate polynomial P with real coefficients and any matrix $K \in \mathbb{R}^{d \times d}$, $P(K)$ shares the same eigenspace with K and its eigenvalues are $\{P(\lambda_i(K))\}_{i \in \{1,\dots,d\}}$. Therefore, the first term on the right-hand side of (5.51) equals

$$\begin{aligned}
\frac{1}{d}\|\widehat{M}^{t'}\Pi\widehat{v}^t\|_2^2 &= \frac{1}{d}\left\|\sum_{i=1}^d \lambda_i(\widehat{M})^{t'}v_i(D)v_i(D)^\top\Pi\widehat{v}^t\right\|_2^2 \\
&= \frac{1}{d}\|\lambda_1(\widehat{M})^{t'}v_1(D)v_1(D)^\top\widehat{v}^t\|_2^2 \\
&= \lambda_1(\widehat{M})^{2t'}\frac{\langle v_1(D), \widehat{v}^t\rangle^2}{d}.
\end{aligned} \tag{5.52}$$

The third term on the right-hand side of (5.51) vanishes:

$$\frac{1}{d}\langle \widehat{M}^{t'}\Pi\widehat{v}^t, \widehat{M}^{t'}\Pi^\perp\widehat{v}^t\rangle$$

$$= \frac{1}{d}\left\langle \lambda_1(\widehat{M})^{t'}\langle v_1(D),\widehat{v}^t\rangle v_1(D), \sum_{i=2}^{d}\lambda_i(\widehat{M})^{t'}\langle v_i(D),\widehat{v}^t\rangle v_i(D)\right\rangle = 0. \quad (5.53)$$

To analyze the second term on the right-hand side of (5.51), we define the matrix

$$\widetilde{M} := \widehat{M}\Pi^\perp = \sum_{i=2}^{d}\lambda_i(\widehat{M})v_i(D)v_i(D)^\top.$$

We then have

$$\frac{1}{d}\|\widehat{M}^{t'}\Pi^\perp\widehat{v}^t\|_2^2 = \frac{1}{d}\left\|\sum_{i=2}^{d}\lambda_i(\widehat{M})^{t'}v_i(D)v_i(D)^\top\widehat{v}^t\right\|_2^2$$

$$= \frac{1}{d}\|\widetilde{M}^{t'}\widehat{v}^t\|_2^2 \leq \frac{\|\widehat{v}^t\|_2^2}{d}\max_{v\in\mathbb{S}^{d-1}}\|\widetilde{M}^{t'}v\|_2^2$$

$$= \frac{\|\widehat{v}^t\|_2^2}{d}\sigma_1(\widetilde{M}^{t'})^2$$

$$= \frac{\|\widehat{v}^t\|_2^2}{d}\lambda_1(\widetilde{M}^{t'})^2 = \frac{\|\widehat{v}^t\|_2^2}{d}\lambda_1(\widetilde{M})^{2t'} = \frac{\|\widehat{v}^t\|_2^2}{d}\lambda_2(\widehat{M})^{2t'},$$

where the passages in the second line follow from the positive definiteness of $\widetilde{M}$.

We have proved in Section 5.4 (see Lemmas 5.4 and 5.5) that almost surely

$$\lim_{d\to\infty}\lambda_2(D) = \lambda_2 := a^\circ\gamma^\circ.$$

Recalling from (3.2) and (3.5) the definitions of ψ, ζ, we can alternatively write $\lambda_2 = \psi(a^\circ) = \zeta(a^\circ)$ as in (3.7). Also recall from (5.42) that $\lambda_1 = a^*\gamma^* = \psi(a^*)$. Under the condition $a^* > a^\circ$, we further have $\lambda_1 = \zeta(a^*)$ as in (3.7). Thus, by the monotonicity of ψ (see Lemma E.1), we obtain the strict inequality $\lambda_1 > \lambda_2$ in the second part of (5.34).

In words, the limiting value of $\lambda_2(D)$ is strictly less than λ_1. In view of (5.43), this gives that $\lim_{d\to\infty}\lambda_2(\widehat{M}) < 1$, which implies

$$\lim_{t'\to\infty}\lim_{t\to\infty}\operatorname*{p-lim\,sup}_{d\to\infty}\frac{1}{d}\|\widehat{M}^{t'}\Pi^\perp\widehat{v}^t\|_2^2$$

$$\leq \lim_{t'\to\infty}\lim_{t\to\infty}\operatorname*{p-lim\,sup}_{d\to\infty}\frac{\|\widehat{v}^t\|_2^2}{d}\lambda_2(\widehat{M})^{2t'}$$

$$\leq \lim_{t'\to\infty}\left(\lim_{t\to\infty}\operatorname*{p-lim}_{d\to\infty}\frac{\|\widehat{v}^t\|_2^2}{d}\right)\left(\lim_{d\to\infty}\lambda_2(\widehat{M})^{2t'}\right) = 0. \quad (5.54)$$

The last equality holds since the limit in the first parentheses is finite (see (5.47)).

Combining (5.52), (5.54), and (5.53), we obtain that the quantity in (5.50) equals

$$\lim_{t' \to \infty} \lim_{t \to \infty} \operatorname*{p-lim}_{d \to \infty} \frac{1}{d} \| \widehat{M}^{t'} \widehat{v}^t \|_2^2$$

$$= \lim_{t' \to \infty} \lim_{t \to \infty} \operatorname*{p-lim}_{d \to \infty} \lambda_1(\widehat{M})^{2t'} \frac{\langle v_1(D), \widehat{v}^t \rangle^2}{d}$$

$$= \lim_{t' \to \infty} \lim_{t \to \infty} \Big(\operatorname*{p-lim}_{d \to \infty} \lambda_1(\widehat{M})^{2t'} \Big) \Big(\operatorname*{p-lim}_{d \to \infty} \frac{\langle v_1(D), \widehat{v}^t \rangle^2}{d} \Big)$$

$$= \Big(\lim_{t' \to \infty} \operatorname*{p-lim}_{d \to \infty} \lambda_1(\widehat{M})^{2t'} \Big) \Big(\lim_{t \to \infty} \operatorname*{p-lim}_{d \to \infty} \frac{\langle v_1(D), \widehat{v}^t \rangle^2}{d} \Big). \qquad (5.55)$$

Now, putting (5.47) and (5.55) together, we arrive at the following relation:

$$\nu^2 = \Big(\lim_{t' \to \infty} \operatorname*{p-lim}_{d \to \infty} \lambda_1(\widehat{M})^{2t'} \Big) \Big(\lim_{t \to \infty} \operatorname*{p-lim}_{d \to \infty} \frac{\langle v_1(D), \widehat{v}^t \rangle^2}{d} \Big).$$

By (5.47), this is equivalent to

$$1 = \Big(\lim_{t' \to \infty} \operatorname*{p-lim}_{d \to \infty} \lambda_1(\widehat{M})^{2t'} \Big) \Big(\lim_{t \to \infty} \operatorname*{p-lim}_{d \to \infty} \frac{\langle v_1(D), \widehat{v}^t \rangle^2}{\|\widehat{v}^t\|_2^2} \Big). \qquad (5.56)$$

This allows us to conclude:

$$\operatorname*{p-lim}_{d \to \infty} \lambda_1(\widehat{M}) = 1, \quad \lim_{t \to \infty} \operatorname*{p-lim}_{d \to \infty} \frac{\langle v_1(D), \widehat{v}^t \rangle^2}{\|\widehat{v}^t\|_2^2} = 1. \qquad (5.57)$$

Indeed, otherwise if the limit of $\lambda_1(\widehat{M})^2$ is different from 1, the right-hand side of (5.56) will either be 0 (if $\operatorname{p-lim}_{d \to \infty} \lambda_1(\widehat{M})^2 \in [0, 1)$) or ∞ (if $\operatorname{p-lim}_{d \to \infty} \lambda_1(\widehat{M})^2 \in (1, \infty)$) once the limit with respect to $t' \to \infty$ is taken. However, this contradicts the left-hand side of (5.56). Since $\widehat{M}$ is positive definite, $\lambda_1(\widehat{M})$ must converge to 1 (instead of -1). Finally, note that by (5.43), the first identity in (5.57) gives that $\operatorname{p-lim}_{d \to \infty} \lambda_1(D) = \lambda_1$ and the second equation says that $\widehat{v}^t$ is asymptotically aligned with $v_1(D)$. This concludes the proof of (5.34). ∎

Proof of (3.12). Since $\widehat{v}^t$ is asymptotically aligned with $v_1(D)$ by (5.34), the overlap between $v_1(D)$ and β^* is the same as that between $\widehat{v}^t$ and β^* in the large t limit. Specifically,

$$\frac{\langle v_1(D), \beta^* \rangle^2}{\|\beta^*\|_2^2} = \Big\langle \frac{\widehat{v}^t}{\|\widehat{v}^t\|_2}, \frac{\beta^*}{\sqrt{d}} \Big\rangle^2 + \Big\langle v_1(D) - \frac{\widehat{v}^t}{\|\widehat{v}^t\|_2}, \frac{\beta^*}{\sqrt{d}} \Big\rangle^2$$

$$+ 2 \Big\langle \frac{\widehat{v}^t}{\|\widehat{v}^t\|_2}, \frac{\beta^*}{\sqrt{d}} \Big\rangle \Big\langle v_1(D) - \frac{\widehat{v}^t}{\|\widehat{v}^t\|_2}, \frac{\beta^*}{\sqrt{d}} \Big\rangle. \qquad (5.58)$$

Note that (5.34) implies

$$\lim_{t\to\infty} \text{p-lim}_{d\to\infty} \left\| \frac{\widehat{v}^t}{\|\widehat{v}^t\|_2} - v_1(D) \right\|_2^2 = 0.$$

Therefore, we have

$$0 \le \lim_{t\to\infty} \text{p-lim}_{d\to\infty} \left\langle v_1(D) - \frac{\widehat{v}^t}{\|\widehat{v}^t\|_2}, \frac{\beta^*}{\sqrt{d}} \right\rangle^2$$

$$\le \lim_{t\to\infty} \text{p-lim}_{d\to\infty} \left\| v_1(D) - \frac{\widehat{v}^t}{\|\widehat{v}^t\|_2} \right\|_2^2 = 0,$$

and

$$0 \le \lim_{t\to\infty} \text{p-lim}_{d\to\infty} \left| \left\langle \frac{\widehat{v}^t}{\|\widehat{v}^t\|_2}, \frac{\beta^*}{\sqrt{d}} \right\rangle \left\langle v_1(D) - \frac{\widehat{v}^t}{\|\widehat{v}^t\|_2}, \frac{\beta^*}{\sqrt{d}} \right\rangle \right|$$

$$\le \lim_{t\to\infty} \text{p-lim}_{d\to\infty} \left\| v_1(D) - \frac{\widehat{v}^t}{\|\widehat{v}^t\|_2} \right\|_2 = 0.$$

Then, taking the limit with respect to d and t on both sides of (5.58), we obtain

$$\text{p-lim}_{d\to\infty} \frac{\langle v_1(D), \beta^* \rangle^2}{\|\beta^*\|_2^2} = \lim_{t\to\infty} \text{p-lim}_{d\to\infty} \frac{\langle \widehat{v}^t, \beta^* \rangle^2}{\|\widehat{v}^t\|_2^2 \cdot d}$$

$$= \frac{\lim_{t\to\infty} \text{p-lim}_{d\to\infty} \frac{1}{d^2} \langle \widehat{v}^t, \beta^* \rangle^2}{\lim_{t\to\infty} \text{p-lim}_{d\to\infty} \frac{1}{d} \|\widehat{v}^t\|_2^2},$$

the right-hand side of which we compute below.

Note that the denominator has already been computed in (5.47) and equals ν^2. The numerator can be computed in a similar way using state evolution. Recalling from (5.41) and (5.40) that $\widehat{v}^t = \Sigma^{-1/2} B v^t$, we have

$$\lim_{t\to\infty} \text{p-lim}_{d\to\infty} \frac{\langle \widehat{v}^t, \beta^* \rangle^2}{d^2}$$

$$= \lim_{t\to\infty} \lim_{d\to\infty} \frac{1}{d^2} \mathbb{E}\big[\mathfrak{B}^{*\top} \Sigma^{-1/2} B V_t \big]^2$$

$$= \lim_{t\to\infty} \chi_t^2 \lim_{d\to\infty} \frac{1}{d^2} \mathbb{E}\big[\mathfrak{B}^{*\top} \Sigma^{-1/2} B \widetilde{\mathfrak{B}}^* \big]^2$$

$$= \Big(\lim_{t\to\infty} \chi_t^2 \Big)\Big(\lim_{d\to\infty} \frac{1}{d^2} \mathbb{E}\big[\mathfrak{B}^{*\top} \Sigma^{-1/2} (\gamma^* I_d - c\Sigma)^{-1} \Sigma \Sigma^{1/2} \mathfrak{B}^* \big]^2 \Big)$$

$$= \chi^2 \mathbb{E}\left[\frac{\overline{\Sigma}}{\gamma^* - c\overline{\Sigma}} \right]^2.$$

Finally, recalling the expressions of χ, σ_V in (5.26), we obtain

$$
\begin{aligned}
\underset{d \to \infty}{\text{p-lim}} \frac{\langle v_1(D), \beta^* \rangle^2}{\|\beta^*\|_2^2} &= \frac{\chi^2 \mathbb{E}[\bar{\Sigma}/(\gamma^* - c\bar{\Sigma})]^2}{v^2} \\
&= \frac{\chi^2 \mathbb{E}[\bar{\Sigma}/(\gamma^* - c\bar{\Sigma})]^2}{\mathbb{E}[\bar{\Sigma}^2/(\gamma^* - c\bar{\Sigma})^2]\chi^2 + \mathbb{E}[\bar{\Sigma}/(\gamma^* - c\bar{\Sigma})^2]\sigma_V^2} \\
&= \frac{(1 - w_2)\mathbb{E}[\bar{\Sigma}/(\gamma^* - c\bar{\Sigma})]^2}{(1 - w_2)\mathbb{E}[\bar{\Sigma}^2/(\gamma^* - c\bar{\Sigma})^2] + w_1\mathbb{E}[\bar{\Sigma}/(\gamma^* - c\bar{\Sigma})^2]} \\
&= \eta^2,
\end{aligned}
$$

as defined in (3.8). ∎

6. Discussion

Information-theoretic limits. In some settings (e.g., phase retrieval), spectral estimators saturate information-theoretic limits when the design matrix is either i.i.d. Gaussian [62] or obtained from a uniformly random orthogonal matrix [35]. That is, below the optimal spectral threshold, *no estimator* can achieve weak recovery, i.e., strictly positive asymptotic overlap with β^*. Thus, it is natural to ask whether the spectral threshold in (3.13) is information-theoretically optimal for weak recovery in problems such as phase retrieval with correlated design. Positive evidence in this regard comes from the comparison with [60] which heuristically derives the information-theoretic weak recovery threshold for general right rotationally invariant designs. As mentioned in Section 3.2, by taking a Gaussian prior on β^*, the model in (1.1) is equivalent to one in which X is right rotationally invariant, and the threshold derived in [60] in fact coincides with the expression in (3.13) (see Remark G.1). An interesting future direction would be to establish whether (and under what conditions) spectral estimators achieve the information-theoretic weak recovery threshold, or conversely to provide evidence of the existence of a statistical-to-computational gap.

Optimal covariance design. Since our results characterize the performance of spectral estimators for a Gaussian design with any covariance Σ, a natural question is to characterize the Σ that induces the maximal overlap. A similar problem is considered in [58] which studies the impact of the spectrum of a bi-rotationally invariant design matrix on the performance of a family of algorithms known as expectation propagation. In contrast, we consider spectral estimators, and our general Gaussian design is only left rotationally invariant. In our context, given the characterization of the limiting overlap $\eta = \eta(\delta, \bar{\Sigma}, \mathcal{T})$ in (3.8) and the expression for the optimal preprocessing $\mathcal{T}^*$ in (3.14), the problem can be formulated as maximizing $\eta(\bar{\Sigma}, \mathcal{T}^*, \delta)$

over $\bar{\Sigma}$, for any fixed δ. Remarkably, Figure 3 in Section 4 shows that picking $\Sigma = I_d$ may not be optimal for the phase retrieval problem. This is in contrast with [58], where it is proved that "spikier" spectra are better for phase retrieval.

Unknown link function. The optimal preprocessing function $\mathcal{T}^*$ in (3.14) depends on the link function q. We now discuss the scenario where q is not exactly known.

In the special case where the link function is parametrized by θ (of fixed dimension) that can be obtained from the moments of the random variable $\bar{Y}$, we can consistently estimate θ with $o(n)$ samples using the empirical moments of the observation vector y. This is, for example, the case when the observations have additive Gaussian noise of unknown variance. One can then apply our spectral estimator using the remaining $n - o(n)$ samples with $\mathcal{T}^*$ constructed from the consistent estimate of q above. By a simple matrix perturbation argument, the same recovery guarantees of the paper continue to hold under the same asymptotic aspect ratio δ.

If q belongs to a non-parametric function class or if the parameters θ cannot be estimated from moments of $\bar{Y}$, then one can still construct $\widetilde{\mathcal{T}}$ (without knowing q) such that the spectral estimator achieves positive asymptotic overlap with β^* when δ is sufficiently large, see [25, Corollary 4.4]. We note that the analysis of [25] considers identity covariance, but as the focus is not on obtaining a tight result in terms of δ, we expect the same to hold if the covariance is well conditioned. Importantly, this comes at the price that $\widetilde{\mathcal{T}}$ no longer achieves the optimal spectral threshold as $\mathcal{T}^*$ does in our Theorem 3.2.

Generalized linear models with unknown link function, also known as single-index models, have been studied in the high-dimensional regime under various assumptions on the link [2, 36, 41, 72, 75, 79]. Sawaya et al. [79] recently studied single-index models under assumptions similar to our paper (Gaussian covariates, and the proportional high-dimensional regime $n/d \to \delta$). Their paper suggests the following three-step procedure to estimate both the signal β^* and the non-parametric link q:

(i)   obtain a pilot estimate of β^*, e.g., by using ridge regression,

(ii) use the pilot estimate to obtain an estimate $\widehat{q}$ of the link function, and

(iii) use $\widehat{q}$ to obtain an improved final estimate $\widehat{\beta}$.

However, the theoretical guarantees in [79] rely on sample splitting which is suboptimal and reduces statistical efficiency. (In sample splitting, the n samples are split into two sets, with the first set used for steps (i)–(ii) and the second set used for obtaining the final estimate in step (iii).) If the link-function is parametrized by a low-dimensional θ of fixed dimension k, e.g., the class of cubic B-spline functions [75], the number of samples required for estimating the link (e.g., via gradient descent) could be relatively small, but we expect this would still be a constant fraction of n since a reasonably accurate pilot estimate is required. In summary, developing sample-

efficient spectral estimators for GLMs with unknown link function is an important open question, which we leave to future work.

Discovering spikes in random matrices via AMP. Our proof strategy offers a new, general methodology for analyzing large spiked random matrices. We expect this strategy to be useful in a variety of statistical inference problems beyond GLMs with correlated Gaussian designs, including rotationally invariant designs [59], mixtures of GLMs [99], principal component analysis with inhomogeneous noise [47], and the universality of spiked random matrices [34, 90]. For many models, the "null" setting in which no information is present can be understood using tools from random matrix theory. When statistically informative components emerge as spectral outliers, our proof recipe can be carried out – as long as an AMP iteration can be designed to simulate the desired power iteration. Suitably combining the analysis for AMP with the random matrix theory arguments for the bulk then allows one to determine the exact outlier locations and estimation accuracy.

Organization of the appendix. Section A contains the proofs of a number of intermediate results useful to show Theorem 3.1. Section B contains the proof of Theorem 3.2. Section C shows how to remove assumptions (A6) and (A7). Section D states and proves a few useful properties of auxiliary functions and parameters. Section E contains the proof of Lemma 5.5. Section F establishes the performance of the whitened spectral estimator. Section G presents some useful auxiliary results.

A. Details of the proof of Theorem 3.1

A.1. Proof of Proposition 5.1

We start by defining the state evolution random vectors $(U_t, V_{t+1})_{t \geq 0}$ in a different, but equivalent form. Let $U_0 \in \mathbb{R}^n$ be a Gaussian random vector whose joint distribution with G is given by $\begin{bmatrix} G \\ U_0 \end{bmatrix} \sim \mathcal{N}(0_{2n}, \Omega_0 \otimes I_n)$, where $\Omega_0 \in \mathbb{R}^{2 \times 2}$ is defined as

$$\Omega_0 = \begin{bmatrix} \frac{1}{\delta} \mathbb{E}[\bar{\Sigma}] & \lim_{n \to \infty} \frac{1}{n} \mathbb{E}[\langle \widetilde{\mathfrak{B}}^*, f_0(\widetilde{\mathfrak{B}}^*) \rangle] \\ \lim_{n \to \infty} \frac{1}{n} \mathbb{E}[\langle \widetilde{\mathfrak{B}}^*, f_0(\widetilde{\mathfrak{B}}^*) \rangle] & \frac{1}{\delta} \left(\text{p-lim}_{d \to \infty} \frac{1}{\sqrt{d}} \|\widetilde{v}^0\|_2 \right)^2 \end{bmatrix}. \tag{A.1}$$

For each $t \geq 0$, define the random vectors $U_t \in \mathbb{R}^n$ and $V_{t+1} \in \mathbb{R}^d$ such that

$$\begin{bmatrix} G \\ U_t \end{bmatrix} \sim \mathcal{N}(0_{2n}, \Omega_t \otimes I_n), \quad V_{t+1} = \chi_{t+1} \widetilde{\mathfrak{B}}^* + \sigma_{V,t+1} W_{V,t+1}, \tag{A.2}$$

where $W_{V,t+1} \sim \mathcal{N}(0_d, I_d)$ is independent of $\widetilde{\mathfrak{B}}^*$ and $\Omega_t \in \mathbb{R}^{2 \times 2}$, $\chi_{t+1} \in \mathbb{R}$, and $\sigma_{V,t+1} \in \mathbb{R}$ are defined recursively as

$$\Omega_t = \begin{bmatrix} \frac{1}{\delta}\mathbb{E}[\bar{\Sigma}] & \lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\langle\widetilde{\mathfrak{B}}^*, f_t(V_t)\rangle] \\ \lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\langle\widetilde{\mathfrak{B}}^*, f_t(V_t)\rangle] & \lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\langle f_t(V_t), f_t(V_t)\rangle] \end{bmatrix}, \quad (A.3)$$

$$\chi_{t+1} = \lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\operatorname{div}_G \widetilde{g}_t(U_t, G, \varepsilon)],$$

$$\sigma_{V,t+1}^2 = \lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\langle g_t(U_t; Y), g_t(U_t; Y)\rangle]. \quad (A.4)$$

Here the function $\widetilde{g}_t \colon (\mathbb{R}^n)^3 \to \mathbb{R}^n$ is given by $\widetilde{g}_t(U_t, G, \varepsilon) = g_t(U_t; q(G, \varepsilon))$.

We now show that the alternative representations of U_t and χ_{t+1} in (A.1) to (A.4) are equivalent to (5.9) to (5.12).

Proposition A.1. *The random vectors (G, U_t) defined in (A.2) can be alternatively written as*

$$U_t = \mu_t G + \sigma_{U,t} W_{U,t}, \quad (A.5)$$

where $(G, W_{U,t}) \sim \mathcal{N}(0_n, \frac{\mathbb{E}[\bar{\Sigma}]}{\delta} I_n) \otimes \mathcal{N}(0_n, I_n)$; for $t = 0$,

$$\mu_0 = \frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\langle\widetilde{\mathfrak{B}}^*, f_0(\widetilde{\mathfrak{B}}^*)\rangle],$$

$$\sigma_{U,0}^2 = \operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\langle\widetilde{v}^0, \widetilde{v}^0\rangle - \frac{\mathbb{E}[\bar{\Sigma}]}{\delta}\mu_0^2 \quad (A.6)$$

and for $t \geq 1$,

$$\mu_t = \frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\langle\widetilde{\mathfrak{B}}^*, f_t(V_t)\rangle],$$

$$\sigma_{U,t}^2 = \lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\langle f_t(V_t), f_t(V_t)\rangle] - \frac{\mathbb{E}[\bar{\Sigma}]}{\delta}\mu_t^2. \quad (A.7)$$

Furthermore, the scalar χ_{t+1} defined in (A.4) can be alternatively written as

$$\chi_{t+1} = \frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\langle G, g_t(U_t; Y)\rangle] - \mu_t \lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\operatorname{div}_{U_t} g_t(U_t; Y)]. \quad (A.8)$$

Proof. The decomposition of U_t in (A.5) and the expressions of $\mu_t, \sigma_{U,t}$ in (A.6) and (A.7) can be easily obtained from (A.1) and (A.3) using the following elementary property of Gaussian random variables. If

$$(G_1, G_2) \sim \mathcal{N}\left(0_2, \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_{2,2} \end{bmatrix}\right),$$

then their joint law can be realized as

$$(G_1, G_2) \stackrel{\mathrm{d}}{=} \left(G_1, \frac{\sigma_{1,2}}{\sigma_{1,1}} G_1 + \sqrt{\sigma_{2,2} - \frac{\sigma_{1,2}^2}{\sigma_{1,1}}} W \right),$$

where $W \sim \mathcal{N}(0, 1)$ is independent of G_1.

To show (A.8), we use the chain rule and Stein's lemma. We have

$$\chi_{t+1} = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\frac{\partial}{\partial G_i} \tilde{g}_t(U_t, G, \varepsilon)_i \right]$$

$$= \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\frac{\partial}{\partial G_i} g_t(U_t; q(G, \varepsilon))_i \right]$$

$$= \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \left(\mathbb{E}\left[\frac{\partial}{\partial G_i} g_t(\mu_t G + \sigma_{U,t} W_{U,t}; q(G, \varepsilon))_i \right] \right.$$

$$\left. - \mu_t \mathbb{E}\left[\frac{\partial}{\partial U_{t,i}} g_t(U_t; Y)_i \right] \right) \tag{A.9}$$

$$= \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]} \mathbb{E}\left[G_i g_t(\mu_t G + \sigma_{U,t} W_{U,t}; q(G, \varepsilon))_i \right] \right.$$

$$\left. - \mu_t \mathbb{E}\left[\frac{\partial}{\partial U_{t,i}} g_t(U_t; Y)_i \right] \right) \tag{A.10}$$

$$= \frac{\delta}{\mathbb{E}[\bar{\Sigma}]} \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left[\langle G, g_t(U_t; Y) \rangle \right] - \mu_t \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left[\mathrm{div}_{U_t} g_t(U_t; Y) \right].$$

Equation (A.9) follows from the chain rule of derivatives:

$$\frac{\partial}{\partial G_i} g_t(U_t; q(G, \varepsilon))_i = \frac{\partial}{\partial G_i} g_t(\mu_t G + \sigma_{U,t} W_{U,t}; q(G, \varepsilon))_i$$

$$= \mu_t \frac{\partial}{\partial U_{t,i}} g_t(U_t; Y)_i + \frac{\partial}{\partial G_i} g_t(U_t; q(G, \varepsilon))_i.$$

Equation (A.10) is by Stein's lemma, noting that $G \sim \mathcal{N}(0_n, \frac{\mathbb{E}[\bar{\Sigma}]}{\delta} I_n)$. ∎

Next, we show the desired state evolution result.

Proof of Proposition 5.1. Define the rescaled version of $\tilde{X}$ as

$$\check{X} := \sqrt{\frac{n}{n+d}} \tilde{X} \in \mathbb{R}^{n \times d}.$$

Note that each entry of $\check{X}$ is i.i.d. according to $\mathcal{N}(0, 1/(n+d))$ and that

$$g = X\beta^* = \tilde{X}\tilde{\beta}^*.$$

Consider a pair of matrix-valued iterates $p^t \in \mathbb{R}^{n \times 2}$ and $q^t \in \mathbb{R}^{d \times 2}$ defined as

$$p^t = \begin{bmatrix} \breve{u}^t & g \end{bmatrix} \in \mathbb{R}^{n \times 2}, \quad q^t = \begin{bmatrix} \breve{v}^t - \breve{\chi}_{t-1}\widetilde{\beta}^* & 0_d \end{bmatrix} \in \mathbb{R}^{d \times 2}, \tag{A.11}$$

where $(\breve{u}^t, \breve{v}^t, \breve{\chi}_{t-1})_{t \geq 0} \subset \mathbb{R}^{n+d+1}$ will be specified later in (A.24). For $(i, j) \in \{1, \dots, n\} \times \{1, 2\}$, we use $p_j^t \in \mathbb{R}^n$ and $p_{i,j}^t \in \mathbb{R}$ to denote the j-th column and the (i, j)-th entry of the matrix p^t, respectively. Similar notation is used for other matrix-valued iterates. Consider also a pair of denoising functions $\pi_t \colon \mathbb{R}^{d \times 3} \to \mathbb{R}^{d \times 2}$ and $\rho_t \colon \mathbb{R}^{n \times 3} \to \mathbb{R}^{n \times 2}$ defined as

$$\begin{aligned}
\pi_t(q^t; \widetilde{\beta}^*) &= \sqrt{\frac{n+d}{n}} \begin{bmatrix} \breve{f}_t(q_1^t + \breve{\chi}_{t-1}\widetilde{\beta}^*) & \widetilde{\beta}^* \end{bmatrix} \in \mathbb{R}^{d \times 2}, \\
\rho_t(p^t; \varepsilon) &= \begin{bmatrix} \sqrt{\frac{n+d}{n}} \breve{g}_t\big(p_1^t; q(p_2^t; \varepsilon)\big) & 0_n \end{bmatrix} \in \mathbb{R}^{n \times 2},
\end{aligned} \tag{A.12}$$

where $(\breve{f}_t, \breve{g}_t)_{t \geq 0}$ will be specified later in (A.24). We claim that the iteration

$$\begin{aligned}
p^{t+1} &= \breve{X}\widetilde{q}^t - \widetilde{p}^{t-1}\ell_t^\top, \quad \widetilde{p}^t = \rho_t(p^t; \varepsilon), \\
\ell_t &= \frac{1}{n+d} \sum_{i=1}^d \begin{bmatrix} \frac{\partial \pi_t(q^t; \widetilde{\beta}^*)_{i,1}}{\partial q_{i,1}^t} & \frac{\partial \pi_t(q^t; \widetilde{\beta}^*)_{i,1}}{\partial q_{i,2}^t} \\ \frac{\partial \pi_t(q^t; \widetilde{\beta}^*)_{i,2}}{\partial q_{i,1}^t} & \frac{\partial \pi_t(q^t; \widetilde{\beta}^*)_{i,2}}{\partial q_{i,2}^t} \end{bmatrix}, \\
q^{t+1} &= \breve{X}^\top \widetilde{p}^t - \widetilde{q}^{t-1}m_t^\top, \quad \widetilde{q}^t = \pi_t(q^t; \widetilde{\beta}^*), \\
m_t &= \frac{1}{n+d} \sum_{i=1}^n \begin{bmatrix} \frac{\partial \rho_t(p^t; \varepsilon)_{i,1}}{\partial p_{i,1}^t} & \frac{\partial \rho_t(p^t; \varepsilon)_{i,1}}{\partial p_{i,2}^t} \\ \frac{\partial \rho_t(p^t; \varepsilon)_{i,2}}{\partial p_{i,1}^t} & \frac{\partial \rho_t(p^t; \varepsilon)_{i,2}}{\partial p_{i,2}^t} \end{bmatrix},
\end{aligned} \tag{A.13}$$

initialized with $\pi_{-1} = 0$, $\rho_{-1} = 0$ and $p^0 = \begin{bmatrix} \breve{u}^0 & g \end{bmatrix}$, $q^0 = \begin{bmatrix} \breve{v}^0 & 0_d \end{bmatrix}$ (for some $\breve{u}^0 \in \mathbb{R}^n$, $\breve{v}^0 \in \mathbb{R}^d$ to be specified later in (A.24)), is equivalent to the following iteration:

$$\begin{aligned}
\breve{u}^{t+1} &= \widetilde{X}\breve{f}_t(\breve{v}^t) - \breve{b}_t \breve{g}_{t-1}(\breve{u}^{t-1}; y), \quad \breve{b}_t = \frac{1}{n} \sum_{i=1}^d \frac{\partial \breve{f}_t(\breve{v}^t)_i}{\partial \breve{v}_i^t}, \\
\breve{v}^{t+1} &= \widetilde{X}^\top \breve{g}_t(\breve{u}^t; y) - \breve{c}_t \breve{f}_{t-1}(\breve{v}^{t-1}), \quad \breve{c}_t = \frac{1}{n} \sum_{i=1}^n \frac{\partial \breve{g}_t(\breve{u}^t; y)_i}{\partial \breve{u}_i^t},
\end{aligned} \tag{A.14}$$

initialized with $\breve{f}_{-1} = 0$, $\breve{g}_{-1} = 0$ and $\breve{u}^0 \in \mathbb{R}^n$, $\breve{v}^0 \in \mathbb{R}^d$.

Let us verify the equivalence. By the design of the matrix-valued iterates in (A.11) and the matrix-valued denoisers in (A.12), we have

$$\begin{aligned}
\widetilde{p}^t &= \rho_t\big(\begin{bmatrix} \breve{u}^t & g \end{bmatrix}; \varepsilon\big) = \begin{bmatrix} \sqrt{\frac{n+d}{n}} \breve{g}_t(\breve{u}^t; q(g; \varepsilon)) & 0_n \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{n+d}{n}} \breve{g}_t(\breve{u}^t; y) & 0_n \end{bmatrix}, \\
\widetilde{q}^t &= \pi_t\big(\begin{bmatrix} \breve{v}^t - \breve{\chi}_{t-1}\widetilde{\beta}^* & 0_d \end{bmatrix}; \widetilde{\beta}^*\big) = \sqrt{\frac{n+d}{n}} \begin{bmatrix} \breve{f}_t(\breve{v}^t) & \widetilde{\beta}^* \end{bmatrix}.
\end{aligned}$$

Furthermore, by chain rule of derivatives, the matrices ℓ_t, m_t specialize to

$$
\ell_t = \frac{1}{n+d} \sum_{i=1}^{d} \begin{bmatrix} \sqrt{\frac{n+d}{n}} \frac{\partial \breve{f}_t(\breve{v}^t)_i}{\partial \breve{v}_i^t} & 0 \\ 0 & 0 \end{bmatrix}
$$
$$
= \sqrt{\frac{n}{n+d}} \begin{bmatrix} \frac{1}{n} \sum_{i=1}^{d} \frac{\partial \breve{f}_t(\breve{v}^t)_i}{\partial \breve{v}_i^t} & 0 \\ 0 & 0 \end{bmatrix} = \sqrt{\frac{n}{n+d}} \begin{bmatrix} \breve{b}_t & 0 \\ 0 & 0 \end{bmatrix},
$$
$$
m_t = \frac{1}{n+d} \sum_{i=1}^{n} \begin{bmatrix} \sqrt{\frac{n+d}{n}} \frac{\partial \breve{g}_t(\breve{u}^t;y)_i}{\partial \breve{u}_i^t} & \sqrt{\frac{n+d}{n}} \frac{\partial \breve{g}_t(\breve{u}^t;q(g;\varepsilon))_i}{\partial g_i} \\ 0 & 0 \end{bmatrix} = \sqrt{\frac{n}{n+d}} \begin{bmatrix} \breve{c}_t & \breve{\chi}_t \\ 0 & 0 \end{bmatrix}.
$$

Using these expressions, we write the iteration in (A.13) as

$$
\begin{bmatrix} \breve{u}^{t+1} & g \end{bmatrix} = \sqrt{\frac{n+d}{n}} \breve{X} \begin{bmatrix} \breve{f}_t(\breve{v}^t) & \tilde{\beta}^* \end{bmatrix}
$$
$$
- \begin{bmatrix} \sqrt{\frac{n+d}{n}} \breve{g}_{t-1}(\breve{u}^{t-1};y) & 0_n \end{bmatrix} \sqrt{\frac{n}{n+d}} \begin{bmatrix} \breve{b}_t & 0 \\ 0 & 0 \end{bmatrix},
$$
$$
\begin{bmatrix} \breve{v}^{t+1} - \breve{\chi}_t \tilde{\beta}^* & 0_d \end{bmatrix} = \breve{X}^\top \begin{bmatrix} \sqrt{\frac{n+d}{n}} \breve{g}_t(\breve{u}^t;y) & 0_n \end{bmatrix}
$$
$$
- \sqrt{\frac{n+d}{n}} \begin{bmatrix} \breve{f}_{t-1}(\breve{v}^{t-1}) & \tilde{\beta}^* \end{bmatrix} \sqrt{\frac{n}{n+d}} \begin{bmatrix} \breve{c}_t & 0 \\ \breve{\chi}_t & 0 \end{bmatrix}.
$$

Expanding the above equations into vector form and using the relation between $\tilde{X}$ and $\breve{X}$, we obtain

$$
\breve{u}^{t+1} = \tilde{X} \breve{f}_t(\breve{v}^t) - \breve{b}_t \breve{g}_{t-1}(\breve{u}^{t-1};y), \quad g = \tilde{X} \tilde{\beta}^*,
$$
$$
\breve{v}^{t+1} = \tilde{X}^\top \breve{g}_t(\breve{u}^t;y) - \breve{c}_t \breve{f}_{t-1}(\breve{v}^{t-1}),
$$

which matches (A.14) and the definition of g.

The iteration in (A.13) is an instance of the abstract graph-based AMP iteration proposed in [42]. To see this, consider a simple graph on two vertices v_1, v_2 with two directed edges $\vec{e} = (v_1, v_2)$ to $\bar{e} = (v_2, v_1)$ between them. The tuple $(\tilde{X}, p^t, \pi_t)$ is associated with the edge $\vec{e}$ and the tuple $(\tilde{X}^\top, q^t, \rho_t)$ is associated with $\bar{e}$. We record below the state evolution results in [42, Section 3.3] for our special case of (A.13), and then translate them to (A.14). For each $t \geq 1$, define two sequences of random matrices

$$
(P_0, P_1, \ldots, P_t) \sim \mathcal{N}(0_{2n(t+1)}, \Theta_t \otimes I_n),
$$
$$
(Q_0, Q_1, \ldots, Q_t) \sim \mathcal{N}(0_{2d(t+1)}, \Xi_t \otimes I_d),
$$

where $P_r \in \mathbb{R}^{n\times 2}$, $Q_r \in \mathbb{R}^{d\times 2}$ ($0 \le r \le t$), and the entries of the covariance matrices $\Theta_t, \Xi_t \in \mathbb{R}^{2(t+1)\times 2(t+1)}$ are specified recursively as follows: for $0 \le r, s \le t$,

$$(\Theta_t)_{r+1,s+1} = \lim_{n\to\infty} \frac{1}{n+d} \mathbb{E}[\pi_r(Q_r; \widetilde{\mathfrak{B}}^*)^\top \pi_s(Q_s \widetilde{\mathfrak{B}}^*)] \in \mathbb{R}^{2\times 2},$$

$$(\Xi_t)_{r+1,s+1} = \lim_{n\to\infty} \frac{1}{n+d} \mathbb{E}[\rho_r(P_r; \varepsilon)^\top \rho_s(P_s; \varepsilon)] \in \mathbb{R}^{2\times 2}.$$

The notation $(P_0, P_1, \ldots, P_t) \in (\mathbb{R}^{n\times 2})^{t+1}$ should be interpreted as a $2n(t + 1)$-dimensional vector given by

$$\begin{bmatrix} (P_0)_1 \\ (P_0)_2 \\ \vdots \\ (P_t)_1 \\ (P_t)_2 \end{bmatrix},$$

where $(P_r)_j$ ($0 \le r \le t$, $j \in \{1, 2\}$) denotes the j-th column of $P_r \in \mathbb{R}^{n\times 2}$. The notation $(Q_0, Q_1, \ldots, Q_t) \in (\mathbb{R}^{d\times 2})^{t+1}$ should be interpreted in a similar way. Accordingly, $\Theta_t, \Xi_t \in \mathbb{R}^{2(t+1)\times 2(t+1)}$ are block matrices whose $(r + 1, s + 1)$-th ($0 \le r, s \le t$) block has size 2×2.

The state evolution result in [42, Theorem 1 and Section 3.3] asserts that for any uniformly pseudo-Lipschitz functions $h_1: \mathbb{R}^{2n(t+1)} \to \mathbb{R}$, $h_2: \mathbb{R}^{2d(t+1)} \to \mathbb{R}$ of finite order,

$$\begin{aligned} \underset{n\to\infty}{\text{p-lim}}\, h_1(p^0, p^1, \ldots, p^t) - \mathbb{E}[h_1(P_0, P_1, \ldots, P_t)] &= 0, \\ \underset{d\to\infty}{\text{p-lim}}\, h_2(q^0, q^1, \ldots, q^t) - \mathbb{E}[h_2(Q_0, Q_1, \ldots, Q_t)] &= 0. \end{aligned} \tag{A.15}$$

With the reduction in (A.11) and (A.12), the state evolution iterates become

$$P_t = \begin{bmatrix} \breve{U}_t & G \end{bmatrix}, \quad Q_t = \begin{bmatrix} \breve{V}_t - \breve{\chi}_{t-1}\widetilde{\mathfrak{B}}^* & 0_d \end{bmatrix},$$

whose covariance structure specializes to

$$\begin{aligned} (\Theta_t)_{r+1,s+1} &= \lim_{n\to\infty} \frac{1}{n+d} \mathbb{E}\left[\frac{n+d}{n} \begin{bmatrix} \breve{f}_r(\breve{V}_r) & \tilde{\beta}^* \end{bmatrix}^\top \begin{bmatrix} \breve{f}_s(\breve{V}_s) & \tilde{\beta}^* \end{bmatrix} \right] \\ &= \begin{bmatrix} \lim_{n\to\infty} \frac{1}{n} \mathbb{E}[\breve{f}_r(\breve{V}_r)^\top \breve{f}_s(\breve{V}_s)] & \lim_{n\to\infty} \frac{1}{n} \mathbb{E}[\breve{f}_r(\breve{V}_r)^\top \widetilde{\mathfrak{B}}^*] \\ \lim_{n\to\infty} \frac{1}{n} \mathbb{E}[\breve{f}_s(\breve{V}_s)^\top \widetilde{\mathfrak{B}}^*] & \lim_{n\to\infty} \frac{1}{n} \mathbb{E}[(\widetilde{\mathfrak{B}}^*)^\top \widetilde{\mathfrak{B}}^*] \end{bmatrix}, \end{aligned} \tag{A.16}$$

$$\begin{aligned} (\Xi_t)_{r+1,s+1} &= \lim_{n\to\infty} \frac{1}{n+d} \mathbb{E}\left[\begin{bmatrix} \sqrt{\frac{n+d}{n}} \breve{g}_r(\breve{U}_r; Y) & 0_n \end{bmatrix}^\top \begin{bmatrix} \sqrt{\frac{n+d}{n}} \breve{g}_s(\breve{U}_s; Y) & 0_n \end{bmatrix} \right] \\ &= \begin{bmatrix} \lim_{n\to\infty} \frac{1}{n} \mathbb{E}[\breve{g}_r(\breve{U}_r; Y)^\top \breve{g}_s(\breve{U}_s; Y)] & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned} \tag{A.17}$$

Reorganizing the elements of P_t, Q_t and Θ_t, Ξ_t, we obtain

$$
\begin{aligned}
(G, \breve{U}_0, \dots, \breve{U}_t) &\sim \mathcal{N}(0_{n(t+2)}, \breve{\Theta}_t \otimes I_n), \\
(\breve{V}_0 - \breve{\chi}_{-1}\mathfrak{B}^*, \dots, \breve{V}_t - \breve{\chi}_{t-1}\mathfrak{B}^*) &\sim \mathcal{N}(0_{d(t+1)}, \breve{\Xi}_t \otimes I_d),
\end{aligned}
\tag{A.18}
$$

where the entries of $\breve{\Theta}_t \in \mathbb{R}^{(t+2)\times(t+2)}$ and $\breve{\Xi}_t \in \mathbb{R}^{(t+1)\times(t+1)}$ are obtained as follows from Θ_t and Ξ_t. Recalling that each entry $(\Theta_t)_{r,s}$, $(\Xi_t)_{r,s}$ of Θ_t, Ξ_t, respectively, is itself a 2×2 matrix, we use $((\Theta_t)_{r,s})_{i,j}$, $((\Xi_t)_{r,s})_{i,j}$ to denote the (i,j)-th $(i,j \in \{1,2\})$ entry of $(\Theta_t)_{r,s}$, $(\Xi_t)_{r,s}$, respectively:

$$
\begin{aligned}
(\breve{\Theta}_t)_{1,1} &= ((\Theta_t)_{1,1})_{2,2}, \quad (\breve{\Theta}_t)_{1,s} = ((\Theta_t)_{s-1,s-1})_{1,2}, && 2 \le s \le t+2, \\
(\breve{\Theta}_t)_{r,s} &= (\breve{\Theta}_t)_{s,r} = ((\Theta_t)_{r-1,s-1})_{1,1}, && 2 \le r \le s \le t+2, \\
(\breve{\Xi}_t)_{r,s} &= (\breve{\Xi}_t)_{s,r} = ((\Xi_t)_{r,s})_{1,1}, && 1 \le r \le s \le t+1.
\end{aligned}
$$

We further transform $\breve{\Theta}_t$ by introducing $\breve{\Omega}_t \in \mathbb{R}^{2\times2}$, $\breve{\Phi}_t \in \mathbb{R}^{(t+1)\times(t+1)}$. First, we have $(G, \breve{U}_t) \sim \mathcal{N}(0_2, \breve{\Omega}_t)$, where

$$
\breve{\Omega}_t = \begin{bmatrix} (\breve{\Theta}_t)_{1,1} & (\breve{\Theta}_t)_{1,t+2} \\ (\breve{\Theta}_t)_{1,t+2} & (\breve{\Theta}_t)_{t+2,t+2} \end{bmatrix} \in \mathbb{R}^{2\times2}.
\tag{A.19}
$$

Next, applying the representation in (A.5) to $(G, \breve{U}_t)$, we write

$$
\breve{U}_t = \breve{\mu}_t G + \breve{\sigma}_{U,t} \breve{W}_{U,t}.
$$

Here $\breve{\mu}_t$ can be derived in a way similar to Proposition A.1:

$$
\breve{\mu}_t = \frac{(\breve{\Theta}_t)_{1,t+2}}{(\breve{\Theta}_t)_{1,1}} = \frac{((\Theta_t)_{t+1,t+1})_{1,2}}{(\breve{\Theta}_t)_{1,1}} = \frac{\delta}{\mathbb{E}[\bar{\Sigma}]} \lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\breve{f}_t(\breve{V}_t)^\top \mathfrak{B}^*], \tag{A.20}
$$

where the last equality is obtained by recalling (A.16). Moreover,

$$
(\breve{\sigma}_{U,0}\breve{W}_{U,0}, \dots, \breve{\sigma}_{U,t}\breve{W}_{U,t}) \sim \mathcal{N}(0_{n(t+1)}, \breve{\Phi}_t \otimes I_n)
$$

are jointly Gaussian whose covariance can be derived from $\breve{\Theta}_t$. For any $0 \le r,s \le t$,

$$
(\breve{\Theta}_t)_{r+2,s+2} = \frac{1}{n}\mathbb{E}[\langle \breve{U}_r, \breve{U}_s \rangle] = \breve{\mu}_r \breve{\mu}_s (\breve{\Theta}_t)_{1,1} + \frac{1}{n}\mathbb{E}[\langle \breve{\sigma}_{U,r}\breve{W}_{U,r}, \breve{\sigma}_{U,s}\breve{W}_{U,s}\rangle],
$$

from which we obtain

$$
\begin{aligned}
(\breve{\Phi}_t)_{r+1,s+1} &= \frac{1}{n}\mathbb{E}[\langle \breve{\sigma}_{U,r}\breve{W}_{U,r}, \breve{\sigma}_{U,s}\breve{W}_{U,s}\rangle] = (\breve{\Theta}_t)_{r+2,s+2} - \breve{\mu}_r \breve{\mu}_s (\breve{\Theta}_t)_{1,1} \\
&= ((\Theta_t)_{r+1,s+1})_{1,1} - \frac{((\Theta_t)_{r+1,r+1})_{1,2}((\Theta_t)_{s+1,s+1})_{1,2}}{((\Theta_t)_{1,1})_{2,2}}. \tag{A.21}
\end{aligned}
$$

We claim that the above expression equals

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\big[\langle \breve{f}_r(\breve{V}_r)-\breve{\mu}_r\widetilde{\mathfrak{B}}^*, \breve{f}_s(\breve{V}_s)-\breve{\mu}_s\widetilde{\mathfrak{B}}^*\rangle\big].$$

Indeed,

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\big[\langle \breve{f}_r(\breve{V}_r)-\breve{\mu}_r\widetilde{\mathfrak{B}}^*, \breve{f}_s(\breve{V}_s)-\breve{\mu}_s\widetilde{\mathfrak{B}}^*\rangle\big]$$

$$=\lim_{n\to\infty}\frac{1}{n}\Big(\mathbb{E}\big[\langle \breve{f}_r(\breve{V}_r), \breve{f}_s(\breve{V}_s)\rangle\big]-\breve{\mu}_s\mathbb{E}\big[\langle \breve{f}_r(\breve{V}_r), \widetilde{\mathfrak{B}}^*\rangle\big]$$

$$-\breve{\mu}_r\mathbb{E}\big[\langle \breve{f}_s(\breve{V}_s), \widetilde{\mathfrak{B}}^*\rangle\big]+\breve{\mu}_r\breve{\mu}_s\mathbb{E}\big[\langle \widetilde{\mathfrak{B}}^*, \widetilde{\mathfrak{B}}^*\rangle\big]\Big)$$

$$=\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\big[\langle \breve{f}_r(\breve{V}_r), \breve{f}_s(\breve{V}_s)\rangle\big]$$

$$-\frac{\delta}{\mathbb{E}[\overline{\Sigma}]}\Big(\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\big[\langle \breve{f}_r(\breve{V}_r), \widetilde{\mathfrak{B}}^*\rangle\big]\Big)\Big(\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\big[\langle \breve{f}_s(\breve{V}_s), \widetilde{\mathfrak{B}}^*\rangle\big]\Big),$$

which agrees with (A.21). In the last equality, we use (A.20).

Finally, for $t\geq 0$, let $\breve{\sigma}_{V,t}\breve{W}_{V,t}:=\breve{V}_t-\breve{\chi}_{t-1}\widetilde{\mathfrak{B}}^*$, where $\breve{W}_{V,t}\sim\mathcal{N}(0,1)$ is independent of $\widetilde{\mathfrak{B}}^*$. From (A.18), we have $(\breve{\sigma}_{V,0}\breve{W}_{V,0},\dots,\breve{\sigma}_{V,t}\breve{W}_{V,t})\sim\mathcal{N}(0_{d(t+1)},\breve{\Xi}_t\otimes I_d)$, where $\breve{\Xi}_t$ has entries

$$(\breve{\Xi}_t)_{r+1,s+1}=\big((\Xi_t)_{r+1,s+1}\big)_{1,1}=\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\big[\langle \breve{g}_r(\breve{U}_r;Y), \breve{g}_s(\breve{U}_s;Y)\rangle\big].\quad\text{(A.22)}$$

With $(\breve{\mu}_t,\breve{\sigma}_{U,t})$ (or equivalently $\breve{\Omega}_t$), $\breve{\Phi}_t,\breve{\chi}_{t-1},\breve{\Xi}_t$ at hand, (A.15) naturally translates to the following state evolution result. For any uniformly pseudo-Lipschitz functions $h_1:\mathbb{R}^{n(t+2)}\to\mathbb{R}$, $h_2:\mathbb{R}^{d(t+2)}\to\mathbb{R}$ of finite order,

$$\operatorname*{p\text{-}lim}_{n\to\infty}h_1(g,\breve{u}_0,\dots,\breve{u}_t)-\mathbb{E}\big[h_1(G,\breve{U}_0,\dots,\breve{U}_t)\big]=0,$$

$$\operatorname*{p\text{-}lim}_{d\to\infty}h_2(\widetilde{\beta}^*,\breve{v}_0,\dots,\breve{v}_t)-\mathbb{E}\big[h_2(\widetilde{\mathfrak{B}}^*,\breve{V}_0,\dots,\breve{V}_t)\big]=0.\quad\text{(A.23)}$$

Note that the AMP iteration in (A.14) is almost the same as that in (5.1) albeit with a difference in time indices. Indeed, the following relabeling maps (A.14) to (5.1), precisely

$$\breve{u}^{2t-1}=u^{t-1},\quad \breve{v}^{2t}=v^t,\qquad t\geq 1,$$

$$\breve{u}^0=0_n,\quad \breve{f}_0(\breve{v}^0)=\breve{v}^0,$$

$$\breve{g}_{2t-1}=g_t,\quad \breve{g}_{2t}=0,\quad \breve{f}_{2t-1}=0,\quad \breve{f}_{2t}=f_t,\qquad\text{(A.24)}$$

$$\breve{\chi}_{2t-2}=0,\quad \breve{\chi}_{2t-1}=\chi_t,\qquad t\geq 1,$$

$$\breve{g}_0=0,\quad \breve{f}_0=0,\quad \breve{\chi}_{-1}=0.$$

The change of indices above is similar to that presented in [42, Appendix A].

The change of time index in (A.24) also maps respectively $(\breve{\mu}_{2t-1}, \breve{\sigma}_{U,2t-1})$ (or equivalently $\breve{\Omega}_{2t-1}$), $\breve{\Phi}_{2t-1}$, $\breve{\chi}_{2t-1}$, $\breve{\Xi}_{2t}$ in (A.19) to (A.21), and (A.22) to $(\mu_t, \sigma_{U,t})$ (or equivalently Ω_t), Φ_t, χ_t, Ψ_t in (5.12), (5.14) to (5.16), and (A.3). Thus, the convergence result in (A.23) translates to (5.17), which completes the proof. ■

A.2. Proof of Lemma 5.2

We start by simplifying the recursion in (5.24) using the distributional properties of various random variables/vectors in (2.2), (5.8), and (5.9). First,

$$\mu_t = \frac{\delta}{\mathbb{E}[\bar{\Sigma}]} \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[(\widetilde{\mathfrak{B}}^*)^\top (\gamma_t I_d - c\Sigma)^{-1} \Sigma (\chi_t \widetilde{\mathfrak{B}}^* + \sigma_{V,t} W_{V,t})\big] \qquad \text{(A.25)}$$

$$= \chi_t \frac{\delta}{\mathbb{E}[\bar{\Sigma}]} \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[(\widetilde{\mathfrak{B}}^*)^\top (\gamma_t I_d - c\Sigma)^{-1} \Sigma \widetilde{\mathfrak{B}}^*\big] \qquad \text{(A.26)}$$

$$= \chi_t \frac{\delta}{\mathbb{E}[\bar{\Sigma}]} \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[\mathfrak{B}^{*\top} \Sigma^{1/2} (\gamma_t I_d - c\Sigma)^{-1} \Sigma \Sigma^{1/2} \mathfrak{B}^*\big] \qquad \text{(A.27)}$$

$$= \frac{1}{\mathbb{E}[\bar{\Sigma}]} \mathbb{E}\bigg[\frac{\bar{\Sigma}^2}{\gamma_t - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma}}\bigg] \chi_t. \qquad \text{(A.28)}$$

Equation (A.25) is by the definition of B_t (see (5.22)) and V_t (see (5.9)). Equation (A.26) holds since $W_{V,t}$ is independent of $\widetilde{\mathfrak{B}}^*$. Equation (A.27) is by the definition of $\widetilde{\mathfrak{B}}^*$ (see (5.8)). In (A.28) we use Proposition G.2, the distribution of $\mathfrak{B}^*$ (see (5.8)) and the assumption $d/n \to 1/\delta$.

Second,

$$\sigma_{U,t}^2 = \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[(\chi_t \widetilde{\mathfrak{B}}^* + \sigma_{V,t} W_{V,t})^\top \Sigma (\gamma_t I_d - c\Sigma)^{-2} \Sigma (\chi_t \widetilde{\mathfrak{B}}^* + \sigma_{V,t} W_{V,t})\big]$$

$$\quad - \frac{\mathbb{E}[\bar{\Sigma}]}{\delta} \mu_t^2$$

$$= \chi_t^2 \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[(\mathfrak{B}^*)^\top \Sigma^{1/2} \Sigma (\gamma_t I_d - c\Sigma)^{-2} \Sigma \Sigma^{1/2} \mathfrak{B}^*\big]$$

$$\quad + \sigma_{V,t}^2 \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[W_{V,t}^\top \Sigma (\gamma_t I_d - c\Sigma)^{-2} \Sigma W_{V,t}\big] - \frac{\mathbb{E}[\bar{\Sigma}]}{\delta} \mu_t^2$$

$$= \frac{1}{\delta} \mathbb{E}\bigg[\frac{\bar{\Sigma}^3}{(\gamma_t - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2}\bigg] \chi_t^2$$

$$\quad + \frac{1}{\delta} \mathbb{E}\bigg[\frac{\bar{\Sigma}^2}{(\gamma_t - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2}\bigg] \sigma_{V,t}^2 - \frac{1}{\delta} \mathbb{E}[\bar{\Sigma}] \mu_t^2 \qquad \text{(A.29)}$$

$$
= \frac{1}{\delta} \left(\mathbb{E}\left[\frac{\overline{\Sigma}^3}{(\gamma_t - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma})^2} \right] - \frac{1}{\mathbb{E}[\overline{\Sigma}]} \mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma_t - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma}} \right]^2 \right) \chi_t^2
$$
$$
+ \frac{1}{\delta} \mathbb{E}\left[\frac{\overline{\Sigma}^2}{(\gamma_t - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma})^2} \right] \sigma_{V,t}^2, \tag{A.30}
$$

where we use (A.28) in (A.30).

Third,

$$
\chi_{t+1} = \frac{\delta}{\mathbb{E}[\overline{\Sigma}]} \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left[G^\top \operatorname{diag}(\mathcal{F}_{a^*}(Y))(\mu_t G + \sigma_{U,t} W_{U,t}) \right]
$$
$$
- \mu_t \mathbb{E}\left[\mathcal{F}_{a^*}(\overline{Y}) \right] \tag{A.31}
$$
$$
= \frac{\delta}{\mathbb{E}[\overline{\Sigma}]} \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left[G^\top \operatorname{diag}(\mathcal{F}_{a^*}(Y)) G \right] \mu_t - \mu_t \mathbb{E}\left[\mathcal{F}_{a^*}(\overline{Y}) \right] \tag{A.32}
$$
$$
= \mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\overline{\Sigma}]} \overline{G}^2 - 1 \right) \mathcal{F}_{a^*}(\overline{Y}) \right] \mu_t \tag{A.33}
$$
$$
= \frac{1}{\mathbb{E}[\overline{\Sigma}]} \mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\overline{\Sigma}]} \overline{G}^2 - 1 \right) \mathcal{F}_{a^*}(\overline{Y}) \right] \mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma_t - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma}} \right] \chi_t. \tag{A.34}
$$

Equation (A.31) is by the definition of U_t (see (A.5)). Equation (A.32) holds since $W_{U,t}$ is independent of G, and hence also independent of Y. Equation (A.33) follows since each entry of G and $\mathcal{F}_{a^*}(Y)$ is i.i.d., and hence

$$
\lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left[G^\top \operatorname{diag}(\mathcal{F}_{a^*}(Y)) G \right] = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[G_i^2 \mathcal{F}_{a^*}(Y_i) \right]
$$
$$
= \mathbb{E}\left[\overline{G}^2 \mathcal{F}_{a^*}(\overline{Y}) \right].
$$

Equation (A.34) follows from (A.28).

Fourth,

$$
\sigma_{V,t+1}^2 = \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left[(\mu_t G + \sigma_{U,t} W_{U,t})^\top \operatorname{diag}(\mathcal{F}_{a^*}(Y))^2 (\mu_t G + \sigma_{U,t} W_{U,t}) \right]
$$
$$
= \mu_t^2 \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left[G^\top \operatorname{diag}(\mathcal{F}_{a^*}(Y))^2 G \right]
$$
$$
+ \sigma_{U,t}^2 \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left[W_{U,t}^\top \operatorname{diag}(\mathcal{F}_{a^*}(Y))^2 W_{U,t} \right]
$$
$$
= \mathbb{E}\left[\overline{G}^2 \mathcal{F}_{a^*}(\overline{Y})^2 \right] \mu_t^2 + \mathbb{E}\left[\mathcal{F}_{a^*}(\overline{Y})^2 \right] \sigma_{U,t}^2 \tag{A.35}
$$

$$
= \frac{1}{\mathbb{E}[\overline{\Sigma}]^2} \mathbb{E}[\overline{G}^2 \mathscr{F}_{a^*}(\overline{Y})^2] \mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma_t - \mathbb{E}[\mathscr{F}_{a^*}(\overline{Y})]\overline{\Sigma}}\right]^2 \chi_t^2
$$
$$
+ \frac{\mathbb{E}[\mathscr{F}_{a^*}(\overline{Y})^2]}{\delta}\left(\mathbb{E}\left[\frac{\overline{\Sigma}^3}{(\gamma_t - \mathbb{E}[\mathscr{F}_{a^*}(\overline{Y})]\overline{\Sigma})^2}\right]\right.
$$
$$
- \frac{1}{\mathbb{E}[\overline{\Sigma}]}\mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma_t - \mathbb{E}[\mathscr{F}_{a^*}(\overline{Y})]\overline{\Sigma}}\right]^2\right)\chi_t^2
$$
$$
+ \frac{\mathbb{E}[\mathscr{F}_{a^*}(\overline{Y})^2]}{\delta}\mathbb{E}\left[\frac{\overline{\Sigma}^2}{(\gamma_t - \mathbb{E}[\mathscr{F}_{a^*}(\overline{Y})]\overline{\Sigma})^2}\right]\sigma_{V,t}^2 \tag{A.36}
$$
$$
= \frac{1}{\delta}\left(\frac{1}{\mathbb{E}[\overline{\Sigma}]}\mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\overline{\Sigma}]}\overline{G}^2 - 1\right)\mathscr{F}_{a^*}(\overline{Y})^2\right]\mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma_t - \mathbb{E}[\mathscr{F}_{a^*}(\overline{Y})]\overline{\Sigma}}\right]^2\right.
$$
$$
+ \mathbb{E}[\mathscr{F}_{a^*}(\overline{Y})^2]\mathbb{E}\left[\frac{\overline{\Sigma}^3}{(\gamma_t - \mathbb{E}[\mathscr{F}_{a^*}(\overline{Y})]\overline{\Sigma})^2}\right]\right)\chi_t^2
$$
$$
+ \frac{\mathbb{E}[\mathscr{F}_{a^*}(\overline{Y})^2]}{\delta}\mathbb{E}\left[\frac{\overline{\Sigma}^2}{(\gamma_t - \mathbb{E}[\mathscr{F}_{a^*}(\overline{Y})]\overline{\Sigma})^2}\right]\sigma_{V,t}^2. \tag{A.37}
$$

Equation (A.36) is by (A.28) and (A.30).

Furthermore, the right-hand side of (5.21) equals:

$$
\lim_{d\to\infty}\frac{1}{d}\mathbb{E}[V_{t+1}^\top B_{t+1}^\top B_{t+1}V_{t+1}]
$$
$$
= \lim_{d\to\infty}\frac{1}{d}\mathbb{E}\big[(\chi_{t+1}\widetilde{\mathfrak{B}}^* + \sigma_{V,t+1}W_{V,t+1})^\top
$$
$$
\cdot \Sigma(\gamma_{t+1}I_d - c\Sigma)^{-2}\Sigma(\chi_{t+1}\widetilde{\mathfrak{B}}^* + \sigma_{V,t+1}W_{V,t+1})\big]
$$
$$
= \chi_{t+1}^2 \lim_{d\to\infty}\frac{1}{d}\mathbb{E}\big[\mathfrak{B}^{*\top}\Sigma^{3/2}(\gamma_{t+1}I_d - c\Sigma)^{-2}\Sigma^{3/2}\mathfrak{B}^*\big]
$$
$$
+ \sigma_{V,t+1}^2 \lim_{d\to\infty}\frac{1}{d}\mathbb{E}\big[W_{V,t+1}^\top\Sigma(\gamma_{t+1}I_d - c\Sigma)^{-2}\Sigma W_{V,t+1}\big]
$$
$$
= \chi_{t+1}^2\mathbb{E}\left[\frac{\overline{\Sigma}^3}{(\gamma_{t+1} - \mathbb{E}[\mathscr{F}_{a^*}(\overline{Y})]\overline{\Sigma})^2}\right] + \sigma_{V,t+1}^2\mathbb{E}\left[\frac{\overline{\Sigma}^2}{(\gamma_{t+1} - \mathbb{E}[\mathscr{F}_{a^*}(\overline{Y})]\overline{\Sigma})^2}\right].
$$

We therefore obtain the following more transparent expression for γ_{t+1} (cf. (5.21)):

$$
1 = \chi_{t+1}^2\mathbb{E}\left[\frac{\overline{\Sigma}^3}{(\gamma_{t+1} - \mathbb{E}[\mathscr{F}_{a^*}(\overline{Y})]\overline{\Sigma})^2}\right]
$$
$$
+ \sigma_{V,t+1}^2\mathbb{E}\left[\frac{\overline{\Sigma}^2}{(\gamma_{t+1} - \mathbb{E}[\mathscr{F}_{a^*}(\overline{Y})]\overline{\Sigma})^2}\right], \tag{A.38}
$$

where $\chi_{t+1}, \sigma_{V,t+1}$ are computed via (A.34) and (A.37). Again, using a similar monotonicity argument as that following (3.3), we readily have that the solution to the above equation must exist in $(s(a^*), \infty)$ and is unique (where we use (b) and (c) in (2.5)), and therefore γ_{t+1} is well defined.

Next, we solve the fixed points of the above state evolution recursion. Suppose the state evolution parameters $\mu_t, \sigma_{U,t}, \chi_{t+1}, \sigma_{V,t+1}, \gamma_{t+1}$ converge to $\mu, \sigma_U, \chi, \sigma_V, \gamma$, respectively, as $t \to \infty$. Then the latter quantities satisfy the following set of equations which are obtained by removing the time indices in (A.28), (A.30), (A.34), (A.37), and (A.38):

$$\mu = \frac{1}{\mathbb{E}[\bar{\Sigma}]}\mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma}}\right]\chi, \tag{A.39}$$

$$\sigma_U^2 = \frac{1}{\delta}\left(\mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2}\right] - \frac{1}{\mathbb{E}[\bar{\Sigma}]}\mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma}}\right]^2\right)\chi^2$$
$$+ \frac{1}{\delta}\mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2}\right]\sigma_V^2, \tag{A.40}$$

$$\chi = \frac{1}{\mathbb{E}[\bar{\Sigma}]}\mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1\right)\mathcal{F}_{a^*}(\bar{Y})\right]\mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma}}\right]\chi, \tag{A.41}$$

$$\sigma_V^2 = \frac{1}{\delta}\left(\frac{1}{\mathbb{E}[\bar{\Sigma}]}\mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1\right)\mathcal{F}_{a^*}(\bar{Y})^2\right]\mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma}}\right]^2\right.$$
$$\left.+ \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})^2]\mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2}\right]\right)\chi^2$$
$$+ \frac{\mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})^2]}{\delta}\mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2}\right]\sigma_V^2, \tag{A.42}$$

$$1 = \mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2}\right]\chi^2 + \mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2}\right]\sigma_V^2. \tag{A.43}$$

We observe from (A.41) that a trivial fixed point of χ is $\chi = 0$. This implies, via (A.39), that $\mu = 0$. Equations (A.42) and (A.43) then become

$$\sigma_V^2 = \frac{\mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})^2]}{\delta}\mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2}\right]\sigma_V^2, \quad 1 = \mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2}\right]\sigma_V^2,$$

from which γ and σ_V^2 can be solved. To be specific, γ is the unique solution in $(s(a^*), \infty)$ to:

$$1 = \frac{\mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})^2]}{\delta}\mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2}\right],$$

and σ_V^2 is given by

$$\sigma_V^2 = \frac{1}{\mathbb{E}[\bar{\Sigma}^2/(\gamma - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2]}.$$

Finally, σ_U^2 can be solved using (A.40): $\sigma_U^2 = 1/\delta$.

Now assume $\chi \neq 0$. Equation (A.41) implies

$$1 = \frac{1}{\mathbb{E}[\bar{\Sigma}]}\mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1\right)\mathcal{F}_{a^*}(\bar{Y})\right]\mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\bar{\Sigma}}\right], \tag{A.44}$$

from which γ can be solved: $\gamma = \gamma^*$. Recall that γ^* (together with a^*) is well defined through (5.4) and a^* is taken to be the largest solution.

Given γ, (A.39), (A.40), (A.42), and (A.43) form a linear system with unknowns μ^2, σ_U^2, χ^2, σ_V^2. Combining (A.43) and (A.42) and using the definitions of w_1, w_2, z_1, z_2 in (3.9), (3.10), and (5.25), we obtain

$$\chi^2 = \frac{1 - w_2}{(1 - w_2)z_1 + w_1 z_2}, \quad \sigma_V^2 = \frac{w_1}{(1 - w_2)z_1 + w_1 z_2}. \tag{A.45}$$

Note that the above solution is valid since $1 - w_2, w_1, z_1, z_2$ are all positive, provided $a^* > a^\circ$ (see item (3) in Proposition D.6 and Proposition G.1). According to (A.39) and (A.40), this immediately implies

$$\mu^2 = \frac{1}{\mathbb{E}[\overline{\Sigma}]^2} \mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma}}\right]^2 \frac{1 - w_2}{(1 - w_2)z_1 + w_1 z_2}, \tag{A.46}$$

$$\begin{aligned}
\sigma_U^2 = {} & \frac{1}{\delta}\left(\mathbb{E}\left[\frac{\overline{\Sigma}^3}{(\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma})^2}\right] - \frac{1}{\mathbb{E}[\overline{\Sigma}]}\mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma}}\right]^2\right) \\
& \cdot \frac{1 - w_2}{(1 - w_2)z_1 + w_1 z_2} \\
& + \frac{1}{\delta}\mathbb{E}\left[\frac{\overline{\Sigma}^2}{(\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma})^2}\right]\frac{w_1}{(1 - w_2)z_1 + w_1 z_2} \\
= {} & \frac{1/\delta}{(1 - w_2)z_1 + w_1 z_2}\left(\mathbb{E}\left[\frac{\overline{\Sigma}^3}{(\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma})^2}\right]\right. \\
& - \frac{1}{\mathbb{E}[\overline{\Sigma}]}\mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma}}\right]^2 \\
& + \frac{1}{\mathbb{E}[\overline{\Sigma}]^2}\mathbb{E}\left[\frac{\overline{\Sigma}^2}{(\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma})^2}\right]\mathbb{E}\left[\overline{G}^2 \mathcal{F}_{a^*}(\overline{Y})^2\right] \\
& \left.\cdot \mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma}}\right]^2\right), \tag{A.47}
\end{aligned}$$

where the last equality follows from the definitions of w_1, w_2. This concludes the proof.

A.3. Proof of Lemma 5.3

For each $t \geq 0$, the next value of $(\mu_{t+1}, \sigma_{U,t+1}, \chi_{t+2}, \sigma_{V,t+2}, \gamma_{t+2})$ only depends on the current value of $(\mu_t, \sigma_{U,t}, \chi_{t+1}, \sigma_{V,t+1}, \gamma_{t+1})$. Hence, to show that the state evolution parameters do not change, it suffices to check that $(\mu_0, \sigma_{U,0}, \chi_1, \sigma_{V,1}, \gamma_1)$ coincides with the fixed point $(\mu, \sigma_U, \chi, \sigma_V, \gamma^*)$.

By the construction of the AMP initializer $(\widetilde{u}^{-1}, \widetilde{v}^0) \in \mathbb{R}^n \times \mathbb{R}^d$, we have $\mu_0 = \mu$ (see (5.31)). It is easy to verify that $\sigma_{U,0}$ given by (5.31) coincides with σ_U derived

in (A.47). Indeed,

$$
\begin{aligned}
\sigma_{U,0}^2 &= \frac{1}{\delta}\left(1 - \mathbb{E}[\bar{\Sigma}]\mu^2\right) \\
&= \frac{1}{\delta}\left(1 - \frac{1}{\mathbb{E}[\bar{\Sigma}]}\mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma^* - \mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})]\bar{\Sigma}}\right]^2 \frac{1 - w_2}{(1 - w_2)z_1 + w_1 z_2}\right) \quad\quad \text{(A.48)} \\
&= \frac{1/\delta}{(1 - w_2)z_1 + w_1 z_2}\bigg((1 - w_2)z_1 + w_1 z_2 \\
&\quad\quad - \frac{1}{\mathbb{E}[\bar{\Sigma}]}\mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma^* - \mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})]\bar{\Sigma}}\right]^2 (1 - w_2)\bigg) \\
&= \frac{1/\delta}{(1 - w_2)z_1 + w_1 z_2}\bigg(\mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma^* - \mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2}\right] \\
&\quad\quad - \frac{1}{\mathbb{E}[\bar{\Sigma}]}\mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma^* - \mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})]\bar{\Sigma}}\right]^2 \\
&\quad\quad + \frac{1}{\mathbb{E}[\bar{\Sigma}]^2}\mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma^* - \mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2}\right]\mathbb{E}\left[\bar{G}^2\mathscr{F}_{a^*}(\bar{Y})^2\right] \\
&\quad\quad \cdot \mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma^* - \mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})]\bar{\Sigma}}\right]^2\bigg) \quad\quad\quad\quad\quad\quad\quad\quad \text{(A.49)} \\
&= \sigma_U^2.
\end{aligned}
$$

We use the expression of μ (see (A.46)) in (A.48) and the expressions of w_1, w_2, z_1, z_2 (see (3.9), (3.10), and (5.25)) in (A.49).

We then verify $\chi_1 = \chi$. By (A.33),

$$
\begin{aligned}
\chi_1 &= \mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1\right)\mathscr{F}_{a^*}(\bar{Y})\right]\mu_0 \\
&= \mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1\right)\mathscr{F}_{a^*}(\bar{Y})\right]\frac{1}{\mathbb{E}[\bar{\Sigma}]}\mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma^* - \mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})]\bar{\Sigma}}\right] \\
&\quad\quad \cdot \sqrt{\frac{1 - w_2}{(1 - w_2)z_1 + w_1 z_2}}.
\end{aligned}
$$

Comparing the above expression with χ in (A.45), we see that it suffices to verify

$$
\mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1\right)\mathscr{F}_{a^*}(\bar{Y})\right]\frac{1}{\mathbb{E}[\bar{\Sigma}]}\mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma^* - \mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})]\bar{\Sigma}}\right] = 1,
$$

which is true since the fixed point $\gamma = \gamma^*$ satisfies (A.44).

Next, we show $\sigma_{V,1} = \sigma_V$. Using (A.35), we have

$$
\begin{aligned}
\sigma_{V,1}^2 &= \mathbb{E}\left[\bar{G}^2\mathscr{F}_{a^*}(\bar{Y})^2\right]\mu_0^2 + \mathbb{E}\left[\mathscr{F}_{a^*}(\bar{Y})^2\right]\sigma_{U,0}^2 \\
&= \mathbb{E}\left[\bar{G}^2\mathscr{F}_{a^*}(\bar{Y})^2\right]\mu^2 + \frac{\mathbb{E}\left[\mathscr{F}_{a^*}(\bar{Y})^2\right]}{\delta}\left(1 - \mathbb{E}[\bar{\Sigma}]\mu^2\right)
\end{aligned}
$$

$$
= \frac{\mathbb{E}[\bar{\Sigma}]}{\delta} \mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1\right)\mathcal{F}_{a*}(\bar{Y})^2\right]\mu^2 + \frac{\mathbb{E}[\mathcal{F}_{a*}(\bar{Y})^2]}{\delta}
$$

$$
= \frac{1}{\delta\mathbb{E}[\bar{\Sigma}]} \mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1\right)\mathcal{F}_{a*}(\bar{Y})^2\right]
$$

$$
\cdot \mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma^* - \mathbb{E}[\mathcal{F}_{a*}(\bar{Y})]\bar{\Sigma}}\right]^2 \frac{1 - w_2}{(1 - w_2)z_1 + w_1 z_2} + \frac{\mathbb{E}[\mathcal{F}_{a*}(\bar{Y})^2]}{\delta}
$$

$$
= \frac{1}{(1 - w_2)z_1 + w_1 z_2}\left(\left(w_1 - \frac{\mathbb{E}[\mathcal{F}_{a*}(\bar{Y})^2]}{\delta}z_1\right)(1 - w_2)\right.
$$

$$
\left. + \frac{\mathbb{E}[\mathcal{F}_{a*}(\bar{Y})^2]}{\delta}\left((1 - w_2)z_1 + w_1 z_2\right)\right) \tag{A.50}
$$

$$
= \frac{1}{(1 - w_2)z_1 + w_1 z_2}\left(w_1 - w_1 w_2 + \frac{\mathbb{E}[\mathcal{F}_{a*}(\bar{Y})^2]}{\delta}w_1 z_2\right)
$$

$$
= \frac{w_1}{(1 - w_2)z_1 + w_1 z_2} \tag{A.51}
$$

$$
= \sigma_V^2.
$$

Equation (A.50) is by the definitions of w_1, z_1. Equation (A.51) is by the definitions of w_2, z_2, in particular, $w_2 = (\mathbb{E}[\mathcal{F}_{a*}(\bar{Y})^2]/\delta)z_2$.

Finally, it remains to verify $\gamma_1 = \gamma^*$. By (A.38), γ_1 is the unique solution to

$$
1 = \chi_1^2 \mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma_1 - \mathbb{E}[\mathcal{F}_{a*}(\bar{Y})]\bar{\Sigma})^2}\right] + \sigma_{V,1}^2 \mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma_1 - \mathbb{E}[\mathcal{F}_{a*}(\bar{Y})]\bar{\Sigma})^2}\right]
$$

$$
= \chi^2 \mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma_1 - \mathbb{E}[\mathcal{F}_{a*}(\bar{Y})]\bar{\Sigma})^2}\right] + \sigma_V^2 \mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma_1 - \mathbb{E}[\mathcal{F}_{a*}(\bar{Y})]\bar{\Sigma})^2}\right]
$$

$$
= \frac{1}{(1 - w_2)z_1 + w_1 z_2}\left((1 - w_2)\mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma_1 - \mathbb{E}[\mathcal{F}_{a*}(\bar{Y})]\bar{\Sigma})^2}\right]\right.
$$

$$
\left. + w_1 \mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma_1 - \mathbb{E}[\mathcal{F}_{a*}(\bar{Y})]\bar{\Sigma})^2}\right]\right).
$$

Rearranging the terms, we have

$$
0 = (1 - w_2)\left(z_1 - \mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma_1 - \mathbb{E}[\mathcal{F}_{a*}(\bar{Y})]\bar{\Sigma})^2}\right]\right)
$$

$$
+ w_1\left(z_2 - \mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma_1 - \mathbb{E}[\mathcal{F}_{a*}(\bar{Y})]\bar{\Sigma})^2}\right]\right). \tag{A.52}
$$

We argue that γ_1 has to equal γ^* for the above equation to hold.

Note that both $(1 - w_2)$ and w_1 are strictly positive (provided $a^* > a^\circ$; see Proposition D.6 (3) and Proposition G.1). If $\gamma_1 < \gamma^*$, then by the definitions of z_1, z_2,

$$z_1 < \mathbb{E}\left[\frac{\overline{\Sigma}^3}{(\gamma_1 - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma})^2}\right], \quad z_2 < \mathbb{E}\left[\frac{\overline{\Sigma}^2}{(\gamma_1 - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma})^2}\right],$$

and hence the right-hand side of (A.52) is strictly positive, which is a contradiction. A similar contradiction can be derived if $\gamma_1 > \gamma^*$. Thus, $\gamma_1 = \gamma^*$. This concludes the proof.

A.4. Proof of Lemma 5.4

Lemma A.2. *Consider the matrix D in (1.2). Define another matrix $\check{D}$ as*

$$\check{D} = \Sigma^{1/2}\check{X}^\top \check{T}\check{X}\Sigma^{1/2} \in \mathbb{R}^{d\times d},$$

where $\check{T} \in \mathbb{R}^{(n-1)\times(n-1)}$ is a diagonal matrix satisfying:

$$\lambda_1(T) \geq \lambda_1(\check{T}) \geq \lambda_2(T) \geq \lambda_2(\check{T}) \geq \cdots$$
$$\cdots \geq \lambda_{n-1}(T) \geq \lambda_{n-1}(\check{T}) \geq \lambda_n(T),$$

and $\check{X} \in \mathbb{R}^{(n-1)\times d}$ consists of i.i.d. $\mathcal{N}(0, 1/n)$ entries, independent of $\check{T}$. Then for every $n, d \geq 1$, it holds almost surely that

$$\lambda_3(\check{D}) \leq \lambda_2(D) \leq \lambda_1(\check{D}). \tag{A.53}$$

Proof. Recall $g = \widetilde{X}\widetilde{\beta}^*$ and

$$D = \Sigma^{1/2}\widetilde{X}^\top \operatorname{diag}\left(\mathcal{T}\left(q(\widetilde{X}\Sigma^{1/2}\beta^*, \varepsilon)\right)\right)\widetilde{X}\Sigma^{1/2}$$
$$= \Sigma^{1/2}\widetilde{X}^\top \operatorname{diag}\left(\mathcal{T}\left(q(\widetilde{X}\widetilde{\beta}^*, \varepsilon)\right)\right)\widetilde{X}\Sigma^{1/2}.$$

We can decompose $\widetilde{X}$ into the sum of two pieces: one along the direction of g and the other perpendicular to g. Furthermore, by isotropy of Gaussians (see [67, Lemma 3.1], [93, Lemma 2.1]), the distribution of $\widetilde{X}$ remains unchanged if the perpendicular part is replaced with an i.i.d. copy. Specifically,

$$\widetilde{X} \stackrel{\mathrm{d}}{=} \Pi_g \widetilde{X} + \Pi_g^\perp \widehat{X},$$

where

$$\Pi_g := \frac{1}{\|g\|_2^2} gg^\top, \quad \Pi_g^\perp := I_n - \Pi_g,$$

and $\widehat{X} \in \mathbb{R}^{n \times d}$ is an i.i.d. copy of $\widetilde{X}$. Using the variational representation of eigenvalues, we can bound the second eigenvalue of D by the first eigenvalue of a related matrix in which T and $\widetilde{X}$ are "decoupled". Indeed,

$$
\lambda_2(D) = \min_{\substack{\mathcal{V} \subset \mathbb{R}^d \\ \dim(\mathcal{V})=d-1}} \max_{v \in \mathcal{V} \cap \mathbb{S}^{d-1}} v^\top \Sigma^{1/2} \widetilde{X}^\top T \widetilde{X} \Sigma^{1/2} v \tag{A.54}
$$

$$
\stackrel{\mathrm{d}}{=} \min_{\substack{\mathcal{V} \subset \mathbb{R}^d \\ \dim(\mathcal{V})=d-1}} \max_{v \in \mathcal{V} \cap \mathbb{S}^{d-1}} v^\top \Sigma^{1/2} \left(\Pi_g \widetilde{X} + \Pi_g^\perp \widehat{X}\right)^\top T \left(\Pi_g \widetilde{X} + \Pi_g^\perp \widehat{X}\right) \Sigma^{1/2} v
$$

$$
= \min_{\substack{\mathcal{V} \subset \mathbb{R}^d \\ \dim(\mathcal{V})=d-1}} \max_{v \in \mathcal{V} \cap \mathbb{S}^{d-1}} v^\top \left(\frac{\Sigma^{1/2} \widetilde{X}^\top g}{\|g\|_2} \frac{g^\top}{\|g\|_2} + \Sigma^{1/2} \widehat{X}^\top \Pi_g^\perp\right)
$$

$$
\cdot T \left(\frac{g}{\|g\|_2} \frac{g^\top \widetilde{X} \Sigma^{1/2}}{\|g\|_2} + \Pi_g^\perp \widehat{X} \Sigma^{1/2}\right) v
$$

$$
\leq \max_{\substack{v \in \mathbb{S}^{d-1} \\ \langle v, \Sigma^{1/2} \widetilde{X}^\top g/\|g\|_2 \rangle = 0}} v^\top \left(\frac{\Sigma^{1/2} \widetilde{X}^\top g}{\|g\|_2} \frac{g^\top}{\|g\|_2} + \Sigma^{1/2} \widehat{X}^\top \Pi_g^\perp\right)
$$

$$
\cdot T \left(\frac{g}{\|g\|_2} \frac{g^\top \widetilde{X} \Sigma^{1/2}}{\|g\|_2} + \Pi_g^\perp \widehat{X} \Sigma^{1/2}\right) v \tag{A.55}
$$

$$
\leq \max_{v \in \mathbb{S}^{d-1}} v^\top \left(\Sigma^{1/2} \widehat{X}^\top \Pi_g^\perp\right) T \left(\Pi_g^\perp \widehat{X} \Sigma^{1/2}\right) v
$$

$$
= \lambda_1 \left(\Sigma^{1/2} \widehat{X}^\top \Pi_g^\perp T \Pi_g^\perp \widehat{X} \Sigma^{1/2}\right).
$$

In (A.54) and subsequent steps, the minimization is over all $(d-1)$-dimensional subspaces $\mathcal{V} \subset \mathbb{R}^d$. In (A.55), instead of minimizing over all $(d-1)$-dimensional subspaces, we take a particular one:

$$
\mathcal{V}_0 = \left\{v \in \mathbb{R}^d : \left\langle v, \frac{\Sigma^{1/2} \widetilde{X}^\top g}{\|g\|_2} \right\rangle = 0\right\} \in \mathbb{R}^d.
$$

Writing the eigendecomposition of $\Pi_g^\perp$ as $\Pi_g^\perp = Q(I_n - e_n e_n^\top) Q^\top$ for some $Q \in \mathbb{O}(n)$ and using the left rotational invariance of $\widehat{X}$, we continue as follows:

$$
\lambda_1 \left(\Sigma^{1/2} \widehat{X}^\top \Pi_g^\perp T \Pi_g^\perp \widehat{X} \Sigma^{1/2}\right)
$$

$$
= \lambda_1 \left(\Sigma^{1/2} \widehat{X}^\top Q(I_n - e_n e_n^\top) Q^\top T Q(I_n - e_n e_n^\top) Q^\top \widehat{X} \Sigma^{1/2}\right)
$$

$$
\stackrel{\mathrm{d}}{=} \lambda_1 \left(\Sigma^{1/2} \widehat{X}^\top (I_n - e_n e_n^\top) Q^\top T Q(I_n - e_n e_n^\top) \widehat{X} \Sigma^{1/2}\right)
$$

$$
= \lambda_1 \left(\Sigma^{1/2} \widehat{X}^\top (I_n - e_n e_n^\top) \widetilde{T} (I_n - e_n e_n^\top) \widehat{X} \Sigma^{1/2}\right), \tag{A.56}
$$

where in (A.56) we define $\widetilde{T} := Q^\top T Q$. Although $\widetilde{T}$ is no longer diagonal, we note that it has the same eigenvalues as T, i.e., $\{\mathcal{T}(y_1), \ldots, \mathcal{T}(y_n)\}$.

For convenience of the proceeding calculations, let us write $\hat{X}$ and $\tilde{T}$ in block forms:

$$\hat{X} = \begin{bmatrix} \hat{X}_{-n} \\ x_n^\top \end{bmatrix}, \quad \tilde{T} = \begin{bmatrix} \tilde{T}_{-n} & s \\ s^\top & \tilde{t}_n \end{bmatrix},$$

where $\hat{X}_{-n} \in \mathbb{R}^{(n-1)\times d}$ consist of the first $n-1$ rows of $\hat{X}$; $\tilde{T}_{-n} \in \mathbb{R}^{(n-1)\times(n-1)}$ is the top-left $(n-1)\times(n-1)$-submatrix of $\tilde{T}$ and $\tilde{t}_n \in \mathbb{R}$ is the bottom-right element of $\tilde{T}$. Note that by the Cauchy interlacing theorem, the eigenvalues of $\tilde{T}$ (i.e., the diagonal elements of T) are interlaced with those of $\tilde{T}_{-n}$, i.e.,

$$\lambda_1(\tilde{T}) \geq \lambda_1(\tilde{T}_{-n}) \geq \lambda_2(\tilde{T}) \geq \lambda_2(\tilde{T}_{-n}) \geq \cdots$$
$$\cdots \geq \lambda_{n-1}(\tilde{T}) \geq \lambda_{n-1}(\tilde{T}_{-n}) \geq \lambda_n(\tilde{T}). \tag{A.57}$$

Now, returning to bounding $\lambda_2(D)$:

$$\lambda_1\big(\Sigma^{1/2}\hat{X}^\top(I_n - e_n e_n^\top)\tilde{T}(I_n - e_n e_n^\top)\hat{X}\Sigma^{1/2}\big)$$
$$= \lambda_1\left(\Sigma^{1/2}\hat{X}^\top \begin{bmatrix} \tilde{T}_{-n} & 0_{n-1} \\ 0_{n-1}^\top & 0 \end{bmatrix} \hat{X}\Sigma^{1/2}\right)$$
$$= \lambda_1\left(\Sigma^{1/2} \begin{bmatrix} \hat{X}_{-n}^\top & x_n \end{bmatrix} \begin{bmatrix} \tilde{T}_{-n} & 0_{n-1} \\ 0_{n-1}^\top & 0 \end{bmatrix} \begin{bmatrix} \hat{X}_{-n} \\ x_n^\top \end{bmatrix} \Sigma^{1/2}\right)$$
$$= \lambda_1\big(\Sigma^{1/2}\hat{X}_{-n}^\top \tilde{T}_{-n}\hat{X}_{-n}\Sigma^{1/2}\big)$$
$$\overset{\mathrm{d}}{=} \lambda_1\big(\Sigma^{1/2}\hat{X}_{-n}^\top \mathrm{diag}\big(\lambda_1(\tilde{T}_{-n}),\dots,\lambda_{n-1}(\tilde{T}_{-n})\big)\hat{X}_{-n}\Sigma^{1/2}\big).$$

The last step follows from the left rotational invariance of $\hat{X}_{-n}$. Denoting

$$\check{X} := \hat{X}_{-n} \in \mathbb{R}^{(n-1)\times d}, \quad \check{T} := \mathrm{diag}\big(\lambda_1(\tilde{T}_{-n}),\dots,\lambda_{n-1}(\tilde{T}_{-n})\big) \in \mathbb{R}^{(n-1)\times(n-1)},$$

we obtain the upper bound in (A.53).

We then prove a lower bound on $\lambda_2(D)$, again using the Courant–Fischer theorem. Recall

$$\lambda_2(D) \overset{\mathrm{d}}{=} \min_{\substack{\mathcal{V}\subset\mathbb{R}^d \\ \dim(\mathcal{V})=d-1}} \max_{v\in\mathcal{V}\cap\mathbb{S}^{d-1}} v^\top \left(\frac{\Sigma^{1/2}\tilde{X}^\top g}{\|g\|_2}\frac{g^\top}{\|g\|_2} + \Sigma^{1/2}\hat{X}^\top \Pi_g^\perp\right)$$
$$\cdot T\left(\frac{g}{\|g\|_2}\frac{g^\top \tilde{X}\Sigma^{1/2}}{\|g\|_2} + \Pi_g^\perp \hat{X}\Sigma^{1/2}\right)v.$$

Let $\mathcal{V}^* \subset \mathbb{R}^d$ be a minimizer. Since $\dim(\mathcal{V}^*) = d-1$, it can be written as

$$\mathcal{V}^* = \big\{v \in \mathbb{R}^d : \langle v, v^*\rangle = 0\big\}$$

for a vector $v^* \in \mathbb{S}^{d-1}$. We proceed as follows:

$$
\begin{aligned}
\lambda_2(D) \stackrel{\mathrm{d}}{=}\ & \max_{v \in \mathbb{S}^{d-1}, \langle v, v^* \rangle = 0} v^\top \left(\frac{\Sigma^{1/2} \widetilde{X}^\top g}{\|g\|_2} \frac{g^\top}{\|g\|_2} + \Sigma^{1/2} \widehat{X}^\top \Pi_g^\perp \right) \\
& \cdot T \left(\frac{g}{\|g\|_2} \frac{g^\top \widetilde{X} \Sigma^{1/2}}{\|g\|_2} + \Pi_g^\perp \widehat{X} \Sigma^{1/2} \right) v \\
\geq\ & \max_{\substack{v \in \mathbb{S}^{d-1}, \langle v, v^* \rangle = 0 \\ \langle v, \Sigma^{1/2} \widetilde{X}^\top g / \|g\|_2 \rangle = 0}} v^\top \left(\frac{\Sigma^{1/2} \widetilde{X}^\top g}{\|g\|_2} \frac{g^\top}{\|g\|_2} + \Sigma^{1/2} \widehat{X}^\top \Pi_g^\perp \right) \\
& \cdot T \left(\frac{g}{\|g\|_2} \frac{g^\top \widetilde{X} \Sigma^{1/2}}{\|g\|_2} + \Pi_g^\perp \widehat{X} \Sigma^{1/2} \right) v \\
=\ & \max_{\substack{v \in \mathbb{S}^{d-1}, \langle v, v^* \rangle = 0 \\ \langle v, \Sigma^{1/2} \widetilde{X}^\top g / \|g\|_2 \rangle = 0}} v^\top \left(\Sigma^{1/2} \widehat{X}^\top \Pi_g^\perp \right) T \left(\Pi_g^\perp \widehat{X} \Sigma^{1/2} \right) v \\
=\ & \max_{v \in \mathcal{U}_0 \cap \mathbb{S}^{d-1}} v^\top \left(\Sigma^{1/2} \widehat{X}^\top \Pi_g^\perp \right) T \left(\Pi_g^\perp \widehat{X} \Sigma^{1/2} \right) v \qquad\qquad \text{(A.58)} \\
\geq\ & \min_{\substack{\mathcal{U} \subset \mathbb{R}^d \\ \dim(\mathcal{U}) = d-2}} \max_{v \in \mathcal{U} \cap \mathbb{S}^{d-1}} v^\top \left(\Sigma^{1/2} \widehat{X}^\top \Pi_g^\perp \right) T \left(\Pi_g^\perp \widehat{X} \Sigma^{1/2} \right) v \\
=\ & \lambda_3 \left(\Sigma^{1/2} \widehat{X}^\top \Pi_g^\perp T \Pi_g^\perp \widehat{X} \Sigma^{1/2} \right).
\end{aligned}
$$

In (A.58), we let

$$
\mathcal{U}_0 := \left\{ v \in \mathbb{R}^d : \langle v, v^* \rangle = \left\langle v, \frac{\Sigma^{1/2} \widetilde{X}^\top g}{\|g\|_2} \right\rangle = 0 \right\} \subset \mathbb{R}^d.
$$

If v^* and $\Sigma^{1/2} \widetilde{X}^\top g / \|g\|_2$ happen to be collinear, then introduce an additional constraint $\langle v, u \rangle = 0$ for an arbitrary vector $u \in \mathbb{S}^{d-1}$ orthogonal to v^* and the '=' in (A.58) becomes '$\geq$'. Furthermore, we have $\dim(\mathcal{U}_0) = d - 2$.

Finally, by the same reasoning as for the upper bound (in particular (A.57)),

$$
\begin{aligned}
& \lambda_3 \left(\Sigma^{1/2} \widehat{X}^\top \Pi_g^\perp T \Pi_g^\perp \widehat{X} \Sigma^{1/2} \right) \\
& \quad \stackrel{\mathrm{d}}{=} \lambda_3 \left(\Sigma^{1/2} \widehat{X}_{-n}^\top \operatorname{diag} \left(\lambda_1(T), \ldots, \lambda_{n-1}(T) \right) \widehat{X}_{-n} \Sigma^{1/2} \right),
\end{aligned}
$$

where $\widehat{X}_{-n} \in \mathbb{R}^{(n-1) \times d}$ has i.i.d. $\mathcal{N}(0, 1/n)$ entries and is independent of everything else. This concludes the proof of Lemma A.2. ∎

Note that (A.57) in the above proof implies that $\check{T}$ has the same limiting spectral distribution as T which is in turn given by $\mathrm{law}(\mathcal{T}(\bar{Y}))$. Now the only difference between the bound in Lemma A.2 and the one in Lemma 5.4 is that n in the latter is replaced with $n - 1$ in the former. However, this is immaterial asymptotically as $n, d \to \infty$ with $n/d \to \delta$.

To prove Lemma 5.4, it then remains to show that both the upper and lower bounds in Lemma A.2 converge to the same limit $\sup \operatorname{supp}(\overline{\mu}_{\widehat{D}})$. It suffices to consider $\lambda_{1,3}(\widehat{D})$ (instead of $\lambda_{1,3}(\check{D})$).

Since the following result may be of independent interest, we isolate the required assumptions and state it in a self-contained manner:

(A4) $n, d \to \infty$ with $n/d \to \delta$.

(A12) $\|\Sigma\|_2$ and $\|T\|_2$ are uniformly bounded over n.

(A13) The empirical spectral distributions μ_T and μ_Σ of T and Σ converge respectively to $\overline{\mu}_T$ and $\overline{\mu}_\Sigma$, with $\overline{\mu}_T, \overline{\mu}_\Sigma \neq \delta_0$. Furthermore, for all $\varsigma > 0$ there exists $n_0 \in \mathbb{N}$ such that whenever $n \geq n_0$ we have

$$\operatorname{supp} \mu_T \subset \operatorname{supp} \overline{\mu}_T + [-\varsigma, \varsigma], \quad \operatorname{supp} \mu_\Sigma \subset \operatorname{supp} \overline{\mu}_\Sigma + [-\varsigma, \varsigma]. \qquad (A.59)$$

(A14) The support of $\overline{\mu}_T$ intersects with $(0, \infty)$, i.e.,

$$\sup \operatorname{supp} \overline{\mu}_T > 0.$$

The uniform boundedness of $\|\Sigma\|_2$ has been assumed in assumption (A2). The uniform boundedness of $\|T\|_2$ follows from the boundedness of $\mathcal{T}$ in assumption (A5). In assumption (A13), the convergence of $\mu_T = \frac{1}{n} \sum_{i=1}^{n} \delta_{\mathcal{T}(q(\langle x_i, \beta^* \rangle, \varepsilon_i))}$ and the first part of (A.59) follows from the law of large numbers; the convergence of μ_Σ has been assumed in assumption (A2) and the second part of (A.59) is the same as (2.1). Neither $\overline{\mu}_T$ nor $\overline{\mu}_\Sigma$ can be δ_0 since $\mathcal{T}$ is not constantly 0 by (2.4), and $\overline{\Sigma}$ is strictly positive. Assumption (A14) is implied by $\sup_{y \in \operatorname{supp}(\overline{Y})} \mathcal{T}(y) > 0$ in assumption (A5).

Lemma A.3 ($\lambda_1(\widehat{D})$ converges to the right edge, [38, Theorem 4.3]). *Suppose that assumptions* (A4) *and* (A12) *to* (A14) *hold true. Consider the matrix $\widehat{D}$ in* (5.32) *and let $\mu_{\widehat{D}}$ denote its empirical spectral distribution. Then, almost surely, $\mu_{\widehat{D}}$ converges to a deterministic probability measure $\overline{\mu}_{\widehat{D}}$ on $\mathbb{R}$ and*

$$\lim_{d \to \infty} \lambda_1(\widehat{D}) = \sup \operatorname{supp}(\overline{\mu}_{\widehat{D}}).$$

Lemma A.4 ($\lambda_3(\widehat{D})$ converges to the right edge). *Suppose that assumptions* (A4) *and* (A12) *to* (A14) *hold true. Then*

$$\lim_{d \to \infty} \lambda_3(\widehat{D}) = \sup \operatorname{supp}(\overline{\mu}_{\widehat{D}}) \quad \text{almost surely}.$$

Proof. To derive the limit, we show a pair of matching upper and lower bounds. Denote $\lambda^\circ = \sup \operatorname{supp}(\overline{\mu}_{\widehat{D}})$. The upper bound is straightforward:

$$\lim_{d \to \infty} \lambda_3(\widehat{D}) \leq \lim_{d \to \infty} \lambda_1(\widehat{D}) = \sup \operatorname{supp}(\overline{\mu}_{\widehat{D}}),$$

where the equality is by Lemma A.3.

As for the lower bound, we would like to show: for any $\lambda < \lambda^\circ$,

$$\lim_{d \to \infty} \lambda_3(\widehat{D}) \geq \lambda \quad \text{almost surely.}$$

By the choice of λ, there exists a constant $c > 0$ such that $\bar{\mu}_{\widehat{D}}(\lambda, \infty) \geq 2c$. Recall that by [98, Theorem 1.2.1], almost surely $\mu_{\widehat{D}}$ weakly converges to $\bar{\mu}_{\widehat{D}}$. Therefore, with probability 1, for every sufficiently large d, $\mu_{\widehat{D}}(\lambda, \infty) \geq c \geq 3/d$. This means

$$\frac{1}{d} \left| \{i \in \{1, \ldots, d\} : \lambda_i(\widehat{D}) \geq \lambda\} \right| \geq \frac{3}{d},$$

that is, $\lambda_3(\widehat{D}) \geq \lambda$, which completes the proof of the lower bound and hence the lemma. ∎

A.5. Proof of (5.49)

Recall from (5.43) and (5.39) the definition of $\widehat{e}^t$. We will first provide a suite of auxiliary bounds on the spectral norms of various matrices in Section A.5.1. They will prove useful in the sequel. We then show in Section A.5.2 that

$$\lim_{t \to \infty} \text{p-}\lim_{n \to \infty} \frac{1}{\sqrt{n}} \|e_1^t\|_2 = 0, \quad \lim_{t \to \infty} \text{p-}\lim_{d \to \infty} \frac{1}{\sqrt{d}} \|e_2^t\|_2 = 0. \tag{A.60}$$

Next, using this, we show in Section A.5.3 that

$$\lim_{t \to \infty} \text{p-}\lim_{d \to \infty} \frac{1}{\sqrt{d}} \|\widehat{e}^t\|_2 = 0. \tag{A.61}$$

Finally, in Section A.5.4 we prove (5.49), i.e.,

$$\lim_{t' \to \infty} \lim_{t \to \infty} \text{p-}\lim_{d \to \infty} \frac{1}{\sqrt{d}} \|\widehat{e}^{t,t'}\|_2 = 0.$$

A.5.1. Bounding the norms of various matrices.
We first recall the following elementary facts regarding the spectral norm, singular values and eigenvalues of a matrix. For any matrix $K \in \mathbb{R}^{n \times d}$,

$$\|K\|_2 = \sigma_1(K) = \sqrt{\lambda_1(K^\top K)} = \sqrt{\lambda_1(K K^\top)}.$$

If K is symmetric ($n = d$), this is further equal to

$$\|K\|_2 = \sqrt{\lambda_1(K^2)} = \max\{|\lambda_1(K)|, |\lambda_n(K)|\}.$$

If K is PSD, then singular values coincide with eigenvalues, and hence

$$\|K\|_2 = \lambda_1(K).$$

Using these facts, we have

$$\lim_{d \to \infty} \|\Sigma\|_2 = \lim_{d \to \infty} \lambda_1(\Sigma) = \sup \mathrm{supp}(\bar{\Sigma}) =: C_\Sigma, \tag{A.62}$$

$$\lim_{n \to \infty} \|T\|_2 = \lim_{d \to \infty} \max_i |\mathcal{T}(y_i)|$$
$$= \max\{|\inf \mathrm{supp}(\mathcal{T}(\bar{Y}))|, |\sup \mathrm{supp}(\mathcal{T}(\bar{Y}))|\} =: C_T, \tag{A.63}$$

$$\lim_{d \to \infty} \|\widetilde{X}\|_2 = \lim_{d \to \infty} \sqrt{\lambda_1(\widetilde{X}^\top \widetilde{X})} = 1 + \frac{1}{\sqrt{\delta}} =: C_{\widetilde{X}}, \tag{A.64}$$

where the last two lines hold almost surely. Note that $C_\Sigma < \infty$ since $\|\Sigma\|_2$ is uniformly bounded (see assumption (A2)) and $C_T < \infty$ since $\mathcal{T}$ is bounded (see assumption (A5)). The last line follows since $\widetilde{X}^\top \widetilde{X}$ is a Wishart matrix and its top eigenvalue converges almost surely to the right edge $(1 + 1/\sqrt{\delta})^2$ of the support of its limiting spectral distribution, the Marchenko–Pastur law [96]. Additionally, note that $\|\Sigma^k\|_2 = C_\Sigma^k$ for any $k \in \mathbb{R}$, since Σ is PSD. Using the sub-multiplicativity of matrix norms, we then have the following bound for D:

$$\lim_{d \to \infty} \|D\|_2 = \lim_{d \to \infty} \|\Sigma^{1/2} \widetilde{X}^\top T \widetilde{X} \Sigma^{1/2}\|_2$$
$$\leq \lim_{d \to \infty} \|\Sigma^{1/2}\|_2^2 \|\widetilde{X}\|_2^2 \|T\|_2 = C_\Sigma C_{\widetilde{X}}^2 C_T =: C_D. \tag{A.65}$$

Since D is a symmetric matrix, $\|D\|_2 = \max\{|\lambda_1(D)|, |\lambda_d(D)|\}$, and therefore for every sufficiently large d, it holds almost surely that

$$-(C_D + 1) \leq \lambda_d(D) \leq \lambda_1(D) \leq C_D + 1. \tag{A.66}$$

The extra $+1$ term is to exclude fluctuation when $d \leq d_0$ for some constant d_0.

Recall that $a^* > \sup \mathrm{supp}(\mathcal{T}(\bar{Y}))$ and denote

$$\check{C}_T := |\inf \mathrm{supp}(\mathcal{T}(\bar{Y}))|, \quad \hat{C}_T := \sup \mathrm{supp}(\mathcal{T}(\bar{Y})) > 0.$$

Then, we have the following bound for F:

$$\lim_{n \to \infty} \|F\|_2 = \lim_{n \to \infty} \max_i \frac{|\mathcal{T}(y_i)|}{a^* - \mathcal{T}(y_i)}$$
$$\leq \lim_{n \to \infty} \frac{\max_i |\mathcal{T}(y_i)|}{a^* - \max_i \mathcal{T}(y_i)} \leq \frac{C_T}{a^* - \hat{C}_T} =: C_F.$$

Recall

$$B = \left(\gamma^* I_d - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\Sigma\right)^{-1} \Sigma,$$

and $\gamma^* > s(a^*)$. Therefore, $\gamma^* I_d - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\Sigma$ is positive definite. We can then bound the spectral norm of B as follows:

$$\lim_{d \to \infty} \|B\|_2 \leq \lim_{d \to \infty} \left\|\gamma^* I_d - \mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})]\Sigma\right\|_2^{-1} \|\Sigma\|_2 \leq \frac{C_\Sigma}{\gamma^* - s(a^*)} =: C_B. \tag{A.67}$$

Recalling $\widetilde{B} = \Sigma^{-1/2}B$ and using (A.62) and (A.67), we have

$$\lim_{d\to\infty} \|\widetilde{B}\|_2 \leq \lim_{d\to\infty} \|\Sigma\|_2^{-1/2}\|B\|_2 \leq \frac{C_B}{\sqrt{\inf \operatorname{supp}(\overline{\Sigma})}} =: C_{\widetilde{B}}.$$

Note that $C_{\widetilde{B}} < \infty$ since $\overline{\Sigma} > 0$ (see assumption (A2)). Recalling $\widehat{M} = \frac{D + \ell I_d}{\lambda_1 + \ell}$ and using (A.65), we have

$$\lim_{d\to\infty} \|\widehat{M}\|_2 \leq \lim_{d\to\infty} \frac{\|D\|_2 + |\ell|}{|\lambda_1 + \ell|} \leq \frac{C_D + |\ell|}{|\lambda_1 + \ell|} =: C_{\widehat{M}}. \tag{A.68}$$

A.5.2. Bounding e_1^t, e_2^t. To prove (A.60), or equivalently,

$$\lim_{t\to\infty} \operatorname*{p-lim}_{n\to\infty} \frac{1}{n}\|e_1^t\|_2^2 = 0, \quad \lim_{t\to\infty} \operatorname*{p-lim}_{d\to\infty} \frac{1}{d}\|e_2^t\|_2^2 = 0,$$

we follow the proof strategy of [63, Lemma 5.3]. The idea is to express these quantities as state evolution parameters and show that they converge to the desired fixed points. Writing

$$\frac{1}{n}\|e_1^t\|_2^2 = \frac{1}{n}\|u^t - u^{t-1}\|_2^2 = \frac{1}{n}\|u^t\|_2^2 + \frac{1}{n}\|u^{t-1}\|_2^2 - \frac{2}{n}\langle u^t, u^{t-1}\rangle,$$

$$\frac{1}{d}\|e_2^t\|_2^2 = \frac{1}{d}\|v^{t+1} - v^t\|_2^2 = \frac{1}{d}\|v^{t+1}\|_2^2 + \frac{1}{d}\|v^t\|_2^2 - \frac{2}{d}\langle v^{t+1}, v^t\rangle,$$

and using the state evolution result in Proposition 5.1, we have

$$\operatorname*{p-lim}_{n\to\infty} \frac{1}{n}\|e_1^t\|_2^2 = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\big[\langle U_t, U_t\rangle\big] + \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\big[\langle U_{t-1}, U_{t-1}\rangle\big]$$

$$- 2\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\big[\langle U_t, U_{t-1}\rangle\big]$$

$$= \frac{\mathbb{E}[\overline{\Sigma}]}{\delta}\mu_t^2 + \sigma_{U,t}^2 + \frac{\mathbb{E}[\overline{\Sigma}]}{\delta}\mu_{t-1}^2 + \sigma_{U,t-1}^2$$

$$- 2\left(\frac{\mathbb{E}[\overline{\Sigma}]}{\delta}\mu_t\mu_{t-1} + \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\big[\langle \sigma_{U,t}W_{U,t}, \sigma_{U,t-1}W_{U,t-1}\rangle\big]\right),$$

and

$$\operatorname*{p-lim}_{d\to\infty} \frac{1}{d}\|e_2^t\|_2^2 = \lim_{d\to\infty} \frac{1}{d}\mathbb{E}\big[\langle V_{t+1}, V_{t+1}\rangle\big] + \lim_{d\to\infty} \frac{1}{d}\mathbb{E}\big[\langle V_t, V_t\rangle\big]$$

$$- 2\lim_{d\to\infty} \frac{1}{d}\mathbb{E}\big[\langle V_{t+1}, V_t\rangle\big]$$

$$= \mathbb{E}[\overline{\Sigma}]\chi_{t+1}^2 + \sigma_{V,t+1}^2 + \mathbb{E}[\overline{\Sigma}]\chi_t^2 + \sigma_{V,t}^2$$

$$- 2\left(\mathbb{E}[\overline{\Sigma}]\chi_{t+1}\chi_t + \lim_{d\to\infty} \frac{1}{d}\mathbb{E}\big[\langle \sigma_{V,t+1}W_{V,t+1}, \sigma_{V,t}W_{V,t}\rangle\big]\right).$$

By Lemma 5.3, the values of $\mu_t, \sigma_{U,t}, \chi_{t+1}, \sigma_{V,t+1}$ do not change with t and are equal to $\mu, \sigma_U, \chi, \sigma_V$. Therefore, to show (A.60), it suffices to show

$$\lim_{t \to \infty} \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[\langle \sigma_{U,t} W_{U,t}, \sigma_{U,t-1} W_{U,t-1} \rangle\big] = \sigma_U^2,$$

$$\lim_{t \to \infty} \lim_{d \to \infty} \frac{1}{d} \mathbb{E}\big[\langle \sigma_{V,t+1} W_{V,t+1}, \sigma_{V,t} W_{V,t} \rangle\big] = \sigma_V^2.$$

From the state evolution, we have

$$
\begin{aligned}
(\Phi_t)_{t+1,t} &= \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[\langle \sigma_{U,t} W_{U,t}, \sigma_{U,t-1} W_{U,t-1} \rangle\big] \\
&= \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[\langle f_t(V_t) - \mu_t \widetilde{\mathfrak{B}}^*, f_{t-1}(V_{t-1}) - \mu_{t-1} \widetilde{\mathfrak{B}}^* \rangle\big] \\
&= \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[\langle f_t(V_t), f_{t-1}(V_{t-1}) \rangle\big] - \mu_t \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[\langle f_{t-1}(V_{t-1}), \widetilde{\mathfrak{B}}^* \rangle\big] \\
&\quad - \mu_{t-1} \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[\langle f_t(V_t), \widetilde{\mathfrak{B}}^* \rangle\big] + \mu_t \mu_{t-1} \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[\langle \widetilde{\mathfrak{B}}^*, \widetilde{\mathfrak{B}}^* \rangle\big] \\
&= \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[\langle f_t(V_t), f_{t-1}(V_{t-1}) \rangle\big] - \frac{\mathbb{E}[\bar{\Sigma}]}{\delta} \mu_t \mu_{t-1}, \qquad\qquad (A.69)
\end{aligned}
$$

where the last equality is by (5.12); and

$$
\begin{aligned}
(\Psi_t)_{t+1,t} &= \lim_{d \to \infty} \frac{1}{d} \mathbb{E}\big[\langle \sigma_{V,t+1} W_{V,t+1}, \sigma_{V,t} W_{V,t} \rangle\big] \\
&= \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[\langle g_t(U_t; Y), g_{t-1}(U_{t-1}; Y) \rangle\big]. \qquad\qquad (A.70)
\end{aligned}
$$

Recall from (5.20) that $g_t(U_t; Y) = F U_t$ and $f_{t+1}(V_{t+1}) = B_{t+1} V_{t+1}$. Therefore, we have

$$
\begin{aligned}
&\lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[\langle f_t(V_t), f_{t-1}(V_{t-1}) \rangle\big] \\
&= \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[\big(\chi_t \widetilde{\mathfrak{B}}^* + \sigma_{V,t} W_{V,t}\big)^\top B_t^\top B_{t-1} \big(\chi_{t-1} \widetilde{\mathfrak{B}}^* + \sigma_{V,t-1} W_{V,t-1}\big)\big] \\
&= \chi_t \chi_{t-1} \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[\mathfrak{B}^{*\top} \Sigma^{1/2} B_t^\top B_{t-1} \Sigma^{1/2} \mathfrak{B}^*\big] \\
&\quad + \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[\big(\sigma_{V,t} W_{V,t}\big)^\top B_t^\top B_{t-1} \big(\sigma_{V,t-1} W_{V,t-1}\big)\big] \\
&= \chi_t \chi_{t-1} \frac{1}{\delta} \mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma_t - c\bar{\Sigma})(\gamma_{t-1} - c\bar{\Sigma})}\right] \\
&\quad + \frac{1}{\delta} \mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma_t - c\bar{\Sigma})(\gamma_{t-1} - c\bar{\Sigma})}\right] \lim_{d \to \infty} \frac{1}{d} \mathbb{E}\big[\langle \sigma_{V,t} W_{V,t}, \sigma_{V,t-1} W_{V,t-1} \rangle\big],
\end{aligned}
$$
$$(A.71)$$

where we use Proposition G.3 in the last step. Similarly, we have

$$
\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\big[\langle g_t(U_t;Y), g_{t-1}(U_{t-1};Y)\rangle\big]
$$

$$
= \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\big[(\mu_t G + \sigma_{U,t} W_{U,t})^\top F^2(\mu_{t-1}G + \sigma_{U,t-1}W_{U,t-1})\big]
$$

$$
= \mu_t\mu_{t-1}\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\big[G^\top F^2 G\big] + \lim_{n\to\infty}\frac{1}{n}\mathbb{E}\big[(\sigma_{U,t}W_{U,t})^\top F^2(\sigma_{U,t-1}W_{U,t-1})\big]
$$

$$
= \mu_t\mu_{t-1}\mathbb{E}\big[\bar{G}^2\mathcal{F}_{a*}(\bar{Y})^2\big]
$$

$$
+ \mathbb{E}\big[\mathcal{F}_{a*}(\bar{Y})^2\big]\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\big[\langle\sigma_{U,t}W_{U,t},\sigma_{U,t-1}W_{U,t-1}\rangle\big]. \tag{A.72}
$$

Letting

$$
\tau_t := \lim_{n\to\infty}\frac{1}{n}\mathbb{E}\big[\langle\sigma_{U,t}W_{U,t},\sigma_{U,t-1}W_{U,t-1}\rangle\big],
$$

$$
\omega_t := \lim_{d\to\infty}\frac{1}{d}\mathbb{E}\big[\langle\sigma_{V,t}W_{V,t},\sigma_{V,t-1}W_{V,t-1}\rangle\big],
$$

and using (A.71) and (A.72) in (A.69) and (A.70), we obtain a pair of recursions for τ_t,ω_t:

$$
\tau_t = \chi_t\chi_{t-1}\frac{1}{\delta}\mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma_t - c\bar{\Sigma})(\gamma_{t-1} - c\bar{\Sigma})}\right]
$$

$$
- \frac{\mathbb{E}[\bar{\Sigma}]}{\delta}\mu_t\mu_{t-1} + \frac{1}{\delta}\mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma_t - c\bar{\Sigma})(\gamma_{t-1} - c\bar{\Sigma})}\right]\omega_t, \tag{A.73}
$$

$$
\omega_{t+1} = \mu_t\mu_{t-1}\mathbb{E}\big[\bar{G}^2\mathcal{F}_{a*}(\bar{Y})^2\big] + \mathbb{E}\big[\mathcal{F}_{a*}(\bar{Y})^2\big]\tau_t. \tag{A.74}
$$

Using (A.73) in (A.74), we further obtain

$$
\omega_{t+1} = \frac{\mathbb{E}[\bar{\Sigma}]}{\delta}\mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1\right)\mathcal{F}_{a*}(\bar{Y})^2\right]\mu_t\mu_{t-1}
$$

$$
+ \chi_t\chi_{t-1}\frac{\mathbb{E}[\mathcal{F}_{a*}(\bar{Y})^2]}{\delta}\mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma_t - c\bar{\Sigma})(\gamma_{t-1} - c\bar{\Sigma})}\right]
$$

$$
+ \frac{\mathbb{E}[\mathcal{F}_{a*}(\bar{Y})^2]}{\delta}\mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma_t - c\bar{\Sigma})(\gamma_{t-1} - c\bar{\Sigma})}\right]\omega_t.
$$

We would like to show

$$
\lim_{t\to\infty}\omega_{t+1} = \sigma_V^2. \tag{A.75}
$$

To this end, we will upper bound the lim sup and lower bound the lim inf both by σ_V^2.

Let

$$
p_t := \frac{\mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})^2]}{\delta} \mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma_t - c\bar{\Sigma})(\gamma_{t-1} - c\bar{\Sigma})} \right],
$$

$$
q_t := \frac{\mathbb{E}[\bar{\Sigma}]}{\delta} \mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]} \bar{G}^2 - 1 \right) \mathscr{F}_{a^*}(\bar{Y})^2 \right] \mu_t \mu_{t-1}
$$

$$
+ \chi_t \chi_{t-1} \frac{\mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})^2]}{\delta} \mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma_t - c\bar{\Sigma})(\gamma_{t-1} - c\bar{\Sigma})} \right],
$$

and

$$
\underline{\omega} = \liminf_{t\to\infty} \omega_{t+1}, \quad \bar{\omega} = \limsup_{t\to\infty} \omega_{t+1}.
$$

Then by subadditivity of $\limsup$,

$$
\bar{\omega} = \limsup_{t\to\infty} q_t + p_t \omega_t
$$

$$
\leq \lim_{t\to\infty} q_t + \left(\lim_{t\to\infty} p_t \right) \left(\limsup_{t\to\infty} \omega_t \right)
$$

$$
= \frac{\mathbb{E}[\bar{\Sigma}]}{\delta} \mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]} \bar{G}^2 - 1 \right) \mathscr{F}_{a^*}(\bar{Y})^2 \right] \mu^2
$$

$$
+ \frac{\mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})^2]}{\delta} \mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma^* - c\bar{\Sigma})^2} \right] \chi^2 + \frac{\mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})^2]}{\delta} \mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma^* - c\bar{\Sigma})^2} \right] \bar{\omega},
$$

where the inequality holds since $\lim_{t\to\infty} p_t \geq 0$. Rearranging terms on both sides gives

$$
\bar{\omega} \leq \left(1 - \frac{\mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})^2]}{\delta} \mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma^* - c\bar{\Sigma})^2} \right] \right)^{-1}
$$

$$
\cdot \left(\frac{\mathbb{E}[\bar{\Sigma}]}{\delta} \mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]} \bar{G}^2 - 1 \right) \mathscr{F}_{a^*}(\bar{Y})^2 \right] \mu^2 + \frac{\mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})^2]}{\delta} \mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma^* - c\bar{\Sigma})^2} \right] \chi^2 \right).
$$

Note that the term in the first parentheses is positive since it is nothing but $1 - w_2$ which is positive whenever $a^* > a^\circ$. We claim that the right-hand side is equal to σ_V^2. This can be seen from the fixed point equations of the state evolution recursion. Indeed, from (A.35) and (A.29), we have the following identity for σ_V^2:

$$
\sigma_V^2 = \mathbb{E}\left[\bar{G}^2 \mathscr{F}_{a^*}(\bar{Y})^2 \right] \mu^2 + \mathbb{E}\left[\mathscr{F}_{a^*}(\bar{Y})^2 \right] \sigma_U^2
$$

$$
= \mathbb{E}\left[\bar{G}^2 \mathscr{F}_{a^*}(\bar{Y})^2 \right] \mu^2 + \frac{\mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})^2]}{\delta} \mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma^* - \mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2} \right] \chi^2
$$

$$
+ \frac{\mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})^2]}{\delta} \mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma^* - \mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})]\bar{\Sigma})^2} \right] \sigma_V^2 - \frac{\mathbb{E}[\mathscr{F}_{a^*}(\bar{Y})^2]}{\delta} \mathbb{E}[\bar{\Sigma}] \mu^2. \quad \text{(A.76)}
$$

Solving for σ_V^2, we obtain exactly the upper bound on $\bar{\omega}$.

Analogously, a lower bound on $\underline{\omega}$ can be derived using superadditivity of $\liminf$:

$$
\begin{aligned}
\underline{\omega} &= \liminf_{t \to \infty} q_t + p_t \omega_t \\
&\geq \lim_{t \to \infty} q_t + \left(\lim_{t \to \infty} p_t \right) \left(\liminf_{t \to \infty} \omega_t \right) \\
&= \frac{\mathbb{E}[\bar{\Sigma}]}{\delta} \mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]} \bar{G}^2 - 1 \right) \mathcal{F}_{a^*}(\bar{Y})^2 \right] \mu^2 \\
&\quad + \frac{\mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})^2]}{\delta} \mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma^* - c\bar{\Sigma})^2} \right] \chi^2 + \frac{\mathbb{E}[\mathcal{F}_{a^*}(\bar{Y})^2]}{\delta} \mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma^* - c\bar{\Sigma})^2} \right] \underline{\omega}.
\end{aligned}
$$

Rearranging and using (A.76) gives $\underline{\omega} \geq \sigma_V^2$. This establishes (A.75).

Next, using (A.75) in (A.73), we get

$$
\lim_{t \to \infty} \tau_t = \frac{1}{\delta} \mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma^* - c\bar{\Sigma})^2} \right] \chi^2 - \frac{\mathbb{E}[\bar{\Sigma}]}{\delta} \mu^2 + \frac{1}{\delta} \mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma^* - c\bar{\Sigma})^2} \right] \sigma_V^2.
$$

By (A.29), the right-hand side is precisely σ_U^2. Therefore, we conclude

$$
\lim_{t \to \infty} \tau_t = \sigma_U^2,
$$

which, together with (A.75), completes the proof of (A.60).

A.5.3. Bounding $\hat{e}^t$. Let us now prove (A.61). Recall from (5.43) and (5.39) that $\hat{e}^t$ comprises the following terms:

$$
\hat{e}^t = \hat{e}_1^t + \hat{e}_2^t + \hat{e}_3^t + \hat{e}_4^t + \hat{e}_5^t + \hat{e}_6^t,
$$

where

$$
\begin{aligned}
\hat{e}_1^t &= \frac{\ell}{\lambda_1 + \ell} \tilde{B} e_2^t + \frac{a^* c}{\lambda_1 + \ell} \Sigma^{1/2} B e_2^t, \\
\hat{e}_2^t &= \frac{a^*}{\lambda_1 + \ell} (b - b_t) \Sigma^{1/2} \tilde{X}^\top F (b_t F + I_n)^{-1} (bF + I_n)^{-1} \tilde{X} B_t v^t, \\
\hat{e}_3^t &= \frac{a^*}{\lambda_1 + \ell} (\gamma^* - \gamma_t) \Sigma^{1/2} \tilde{X}^\top F (bF + I_n)^{-1} \\
&\quad \cdot \tilde{X} (\gamma_t I_d - c\Sigma)^{-1} (\gamma^* I_d - c\Sigma)^{-1} \Sigma v^t, \\
\hat{e}_4^t &= \frac{a^* (c_t - c)}{\lambda_1 + \ell} B_t v^t, \\
\hat{e}_5^t &= \frac{a^* c}{\lambda_1 + \ell} (\gamma_t - \gamma^*) \Sigma^{1/2} (\gamma_t I_d - c\Sigma)^{-1} (\gamma^* I_d - c\Sigma)^{-1} \Sigma v^t, \\
\hat{e}_6^t &= \frac{a^* b_t}{\lambda_1 + \ell} \Sigma^{1/2} \tilde{X}^\top F^2 (b_t F + I_n)^{-1} e_1^t.
\end{aligned}
$$

Since the AMP is initialized so that the state evolution parameters stay fixed (see Lemma 5.3), for every $t \geq 1$ and $\gamma_t = \gamma^*$, we immediately get

$$\hat{e}_3^t = \hat{e}_5^t = 0_d. \tag{A.77}$$

By convergence of the empirical spectral distribution of Σ (see assumption (A2)), for every $t \geq 1$,

$$\lim_{d \to \infty} b_t = \lim_{d \to \infty} \frac{d}{n} \operatorname{Tr}\big((\gamma_t I_d - c\Sigma)^{-1}\Sigma\big) = \frac{1}{\delta}\mathbb{E}\left[\frac{\bar{\Sigma}}{\gamma_t - c\bar{\Sigma}}\right] = b,$$

and consequently

$$\operatorname*{p-lim}_{d \to \infty} \frac{1}{\sqrt{d}}\|\hat{e}_2^t\|_2 = 0. \tag{A.78}$$

By convergence of the noise sequence $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ (see assumption (A3)) and independence of covariate vectors $(x_1, \ldots, x_n)$ (see assumption (A2)),

$$\operatorname*{p-lim}_{n \to \infty} c_t = \operatorname*{p-lim}_{n \to \infty} \frac{1}{n} \operatorname{Tr}(F) = \mathbb{E}\big[\mathscr{F}_{a^*}(\bar{Y})\big] = c,$$

and consequently

$$\operatorname*{p-lim}_{d \to \infty} \frac{1}{\sqrt{d}}\|\hat{e}_4^t\|_2 = 0. \tag{A.79}$$

We use the bounds developed in the previous sections to bound $\hat{e}_1^t$ and $\hat{e}_6^t$. Specifically,

$$\lim_{t \to \infty} \operatorname*{p-lim}_{d \to \infty} \frac{1}{\sqrt{d}}\|\hat{e}_1^t\|_2$$

$$\leq \lim_{t \to \infty} \operatorname*{p-lim}_{d \to \infty} \left|\frac{\ell}{\lambda_1 + \ell}\right|\|\tilde{B}\|_2\frac{\|e_2^t\|_2}{\sqrt{d}} + \left|\frac{a^*c}{\lambda_1 + \ell}\right|\|\Sigma\|_2^{1/2}\|B\|_2\frac{\|e_2^t\|_2}{\sqrt{d}}$$

$$\leq \left(\left|\frac{\ell}{\lambda_1 + \ell}\right|C_{\tilde{B}} + \left|\frac{a^*c}{\lambda_1 + \ell}\right|\sqrt{C_\Sigma}C_B\right)\lim_{t \to \infty} \operatorname*{p-lim}_{d \to \infty} \frac{\|e_2^t\|_2}{\sqrt{d}} = 0, \tag{A.80}$$

$$\lim_{t \to \infty} \operatorname*{p-lim}_{d \to \infty} \frac{1}{\sqrt{d}}\|\hat{e}_6^t\|_2$$

$$\leq \lim_{t \to \infty} \operatorname*{p-lim}_{d \to \infty} \left|\frac{a^*b_t}{\lambda_1 + \ell}\right|\|\Sigma\|_2^{1/2}\|\tilde{X}\|_2\|F\|_2^2\|(b_t F + I_n)^{-1}\|_2\frac{\|e_1^t\|_2}{\sqrt{d}}$$

$$= \lim_{t \to \infty} \operatorname*{p-lim}_{d \to \infty} \left|\frac{a^*b_t}{\lambda_1 + \ell}\right|\|\Sigma\|_2^{1/2}\|\tilde{X}\|_2\|F\|_2^2\|I_n - \frac{T}{a^*}\|_2\frac{\|e_1^t\|_2}{\sqrt{d}} \tag{A.81}$$

$$= \frac{\sqrt{C_\Sigma}C_{\tilde{X}}C_F^2(a^* - \check{C}_T)}{|\lambda_1 + \ell|}\lim_{t \to \infty} \operatorname*{p-lim}_{d \to \infty} \frac{\|e_1^t\|_2}{\sqrt{d}} = 0. \tag{A.82}$$

To obtain (A.81), it is useful to recall $F = T(a^* I_n - T)^{-1}$ (see (5.19)) and observe from (5.23) and (5.4) that $b_t = 1$ for every $t \geq 1$ (where we use $\gamma_t = \gamma^*$ for every $t \geq 1$ from Lemma 5.3).

Combining (A.77) to (A.80) and (A.82) yields (A.61), as required.

A.5.4. Bounding $\widehat{e}^{t,t'}$. Finally, we prove (5.49). Recalling the definition of $\widehat{e}^{t,t'}$ in (5.46) and using the triangle inequality and the sub-multiplicativity of norms, we have

$$
\lim_{t'\to\infty} \lim_{t\to\infty} \operatorname*{p-lim}_{d\to\infty} \frac{1}{\sqrt{d}} \|\widehat{e}^{t,t'}\|_2
$$

$$
\leq \lim_{t'\to\infty} \lim_{t\to\infty} \operatorname*{p-lim}_{d\to\infty} \frac{1}{\sqrt{d}} \sum_{s=1}^{t'} \|\widehat{M}\|_2^{t'-s} \|\widehat{e}^{t+s-1}\|_2
$$

$$
= \lim_{t'\to\infty} \lim_{t\to\infty} \sum_{s=1}^{t'} \Big(\lim_{d\to\infty} \|\widehat{M}\|_2^{t'-s} \Big) \Big(\operatorname*{p-lim}_{d\to\infty} \frac{1}{\sqrt{d}} \|\widehat{e}^{t+s-1}\|_2 \Big)
$$

$$
\leq \lim_{t'\to\infty} \sum_{s=1}^{t'} C_{\widehat{M}}^{t'-s} \Big(\lim_{t\to\infty} \operatorname*{p-lim}_{d\to\infty} \frac{1}{\sqrt{d}} \|\widehat{e}^{t+s-1}\|_2 \Big) = 0,
$$

which implies (5.49). The inequality in the penultimate line is by (A.68) and the last equality is by (A.61).

B. Proof of Theorem 3.2

We first prove item (2) of Theorem 3.2. Suppose that the condition $a^* > a^\circ$ holds for some $\mathcal{T} \in \mathfrak{T}$. If φ is strictly decreasing on $(\sup \operatorname{supp}(\mathcal{T}(\bar{Y})), \infty)$, this condition is equivalent to the following one:

$$
1 < \frac{1}{\mathbb{E}[\bar{\Sigma}]} \mathbb{E}\Big[\Big(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]} \bar{G}^2 - 1 \Big) \mathcal{F}_{a^\circ}(\bar{Y}) \Big] \mathbb{E}\Big[\frac{\bar{\Sigma}^2}{\gamma^\circ - \mathbb{E}[\mathcal{F}_{a^\circ}(\bar{Y})]\bar{\Sigma}} \Big], \tag{B.1}
$$

by item (4) of Proposition D.6. We assume $a^\circ = 1$. This assumption is without loss of generality due to scaling invariance. Indeed, the threshold condition for δ (i.e., (B.1) above) and the self-consistent equations for (a°, γ°) (see (D.15) and Lemma D.5) only depend on $(a^\circ, \mathcal{T})$ through $\mathcal{F}_{a^\circ}(\bar{Y})$. Therefore, they continue to hold if $(a^\circ, \mathcal{T})$ is replaced[4] with $(1, \mathcal{T}/a^\circ)$. Let $\mathcal{J}(y) = \mathcal{T}(y)/(1 - \mathcal{T}(y))$ for notational convenience. The definition of (a°, γ°) in (D.15) can then be written as

$$
1 = \frac{1}{\delta} \mathbb{E}[\mathcal{J}(\bar{Y})^2] \mathbb{E}\Big[\Big(\frac{\bar{\Sigma}}{\gamma^\circ - \mathbb{E}[\mathcal{J}(\bar{Y})]\bar{\Sigma}} \Big)^2 \Big], \quad 1 = \frac{1}{\delta} \mathbb{E}\Big[\frac{\bar{\Sigma}}{\gamma^\circ - \mathbb{E}[\mathcal{J}(\bar{Y})]\bar{\Sigma}} \Big]. \tag{B.2}
$$

---

[4] Note that $a^\circ > \sup \operatorname{supp}(\mathcal{T}(\bar{Y})) > 0$.

Let $p_{\bar{G}}$ denote the density of $\bar{G} \sim \mathcal{N}(0, \mathbb{E}[\bar{\Sigma}]/\delta)$, and $p(\cdot \mid g)$ the conditional density of $y = q(g, \varepsilon) \in \mathbb{R}$ given $g \in \mathbb{R}$ where $\varepsilon \sim P_\varepsilon$. Then, using the Cauchy–Schwarz inequality, the second factor on the right-hand side of (B.1) can be bounded as follows:

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1\right)\mathcal{F}_{a\circ}(\bar{Y})\right] &= \mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1\right)\mathcal{J}(\bar{Y})\right] \\
&= \int_{\mathrm{supp}(\bar{Y})} \int_{\mathbb{R}} p_{\bar{G}}(g)p(y \mid g)\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}g^2 - 1\right)\mathcal{J}(y)\,\mathrm{d}g\,\mathrm{d}y \\
&= \int_{\mathrm{supp}(\bar{Y})} \mathbb{E}\left[p(y \mid \bar{G})\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1\right)\right]\mathcal{J}(y)\,\mathrm{d}y \\
&= \int_{\mathrm{supp}(\bar{Y})} \frac{\mathbb{E}\left[p(y \mid \bar{G})(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1)\right]}{\sqrt{\mathbb{E}[p(y \mid \bar{G})]}} \cdot \sqrt{\mathbb{E}[p(y \mid \bar{G})]}\,\mathcal{J}(y)\,\mathrm{d}y \\
&\leq \left(\int_{\mathrm{supp}(\bar{Y})} \frac{\mathbb{E}\left[p(y \mid \bar{G})(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1)\right]^2}{\mathbb{E}[p(y \mid \bar{G})]}\,\mathrm{d}y\right)^{1/2} \\
&\quad \cdot \left(\int_{\mathrm{supp}(\bar{Y})} \mathbb{E}\left[p(y \mid \bar{G})\right]\mathcal{J}(y)^2\,\mathrm{d}y\right)^{1/2} \\
&= \left(\int_{\mathrm{supp}(\bar{Y})} \frac{\mathbb{E}\left[p(y \mid \bar{G})(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1)\right]^2}{\mathbb{E}[p(y \mid \bar{G})]}\,\mathrm{d}y\right)^{1/2}\mathbb{E}\left[\mathcal{J}(\bar{Y})^2\right]^{1/2}. \quad \text{(B.3)}
\end{aligned}
$$

Applying the Cauchy–Schwarz inequality to the third factor on the right-hand side of (B.1), we obtain

$$
\begin{aligned}
\frac{1}{\mathbb{E}[\bar{\Sigma}]}\mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma^\circ - \mathbb{E}[\mathcal{F}_{a\circ}(\bar{Y})]\bar{\Sigma}}\right] &= \frac{1}{\mathbb{E}[\bar{\Sigma}]}\mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma^\circ - \mathbb{E}[\mathcal{J}(\bar{Y})]\bar{\Sigma}}\right] \\
&= \frac{1}{\mathbb{E}[\bar{\Sigma}]}\mathbb{E}\left[\frac{\bar{\Sigma}}{\gamma^\circ - \mathbb{E}[\mathcal{J}(\bar{Y})]\bar{\Sigma}} \cdot \bar{\Sigma}\right] \\
&\leq \frac{\mathbb{E}[\bar{\Sigma}^2]^{1/2}}{\mathbb{E}[\bar{\Sigma}]}\mathbb{E}\left[\left(\frac{\bar{\Sigma}}{\gamma^\circ - \mathbb{E}[\mathcal{J}(\bar{Y})]\bar{\Sigma}}\right)^2\right]^{1/2}. \quad \text{(B.4)}
\end{aligned}
$$

Combining (B.3) and (B.4), we have that the right-hand side of (B.1) is bounded from above by

$$
\begin{aligned}
&\frac{\mathbb{E}[\bar{\Sigma}^2]^{1/2}}{\mathbb{E}[\bar{\Sigma}]}\left(\int_{\mathrm{supp}(\bar{Y})} \frac{\mathbb{E}\left[p(y \mid \bar{G})(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1)\right]^2}{\mathbb{E}[p(y \mid \bar{G})]}\,\mathrm{d}y\right)^{1/2} \\
&\quad \cdot \mathbb{E}\left[\mathcal{J}(\bar{Y})^2\right]^{1/2}\mathbb{E}\left[\left(\frac{\bar{\Sigma}}{\gamma^\circ - \mathbb{E}[\mathcal{J}(\bar{Y})]\bar{\Sigma}}\right)^2\right]^{1/2} \\
&= \frac{\mathbb{E}[\bar{\Sigma}^2]^{1/2}}{\mathbb{E}[\bar{\Sigma}]}\left(\int_{\mathrm{supp}(\bar{Y})} \frac{\mathbb{E}\left[p(y \mid \bar{G})(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1)\right]^2}{\mathbb{E}[p(y \mid \bar{G})]}\,\mathrm{d}y\right)^{1/2}\sqrt{\delta},
\end{aligned}
$$

where the equality follows from the first identity in (B.2). Using this in (B.1), we have

$$\delta > \frac{\mathbb{E}[\bar{\Sigma}]^2}{\mathbb{E}[\bar{\Sigma}^2]} \left(\int_{\text{supp}(\bar{Y})} \frac{\mathbb{E}\big[p(y \mid \bar{G})\big(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1\big)\big]^2}{\mathbb{E}[p(y \mid \bar{G})]} \, dy \right)^{-1}. \tag{B.5}$$

In words, the condition above (which is independent of the choice of $\mathcal{T}$) holds for any $\mathcal{T}$ that satisfies (B.1) and therefore achieves a positive overlap.

In the following, we show that the condition above is tight by proving item (1) of Theorem 3.2. Specifically, whenever (B.5) holds, we exhibit a preprocessing function $\mathcal{T}^*: \mathbb{R} \to \mathbb{R}$ that meets (B.1) and therefore must induce a positive overlap.

Suppose that (B.5) holds. As before, we choose the scaling such that $a^\circ = 1$. Constructing $\mathcal{T}^*(y)$ is equivalent to constructing

$$\mathcal{J}^*(y) = \frac{\mathcal{T}^*(y)}{1 - \mathcal{T}^*(y)}. \tag{B.6}$$

We require the following notation. Denote the right-hand side of (B.5) by $\Delta(\delta)$. Moreover, we have

$$m_0(y) := \mathbb{E}\big[p(y \mid \bar{G})\big], \quad m_2(y) := \mathbb{E}\Big[p(y \mid \bar{G}) \cdot \frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2\Big].$$

Before presenting the construction of $\mathcal{J}^*$, we first observe that the integrals of both m_0 and m_2 are equal to 1:

$$\int_{\text{supp}(\bar{Y})} m_0(y) \, dy = \mathbb{E}\Big[\int_{\text{supp}(\bar{Y})} p(y \mid \bar{G}) \, dy\Big] = 1,$$

$$\int_{\text{supp}(\bar{Y})} m_2(y) \, dy = \mathbb{E}\Big[\Big(\int_{\text{supp}(\bar{Y})} p(y \mid \bar{G}) \, dy\Big) \frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2\Big] \tag{B.7}$$

$$= \mathbb{E}\Big[\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2\Big] = 1.$$

Now, consider

$$\mathcal{J}^*(y) := \sqrt{\frac{\Delta(\delta)}{\delta}}\Big(\frac{m_2(y)}{m_0(y)} - 1\Big). \tag{B.8}$$

We claim that $\mathcal{J}^*$ satisfies (B.1) and (B.2) and therefore attains positive overlap. In fact, we claim that $\mathcal{J}^*$ satisfies a stronger condition than (B.1), which is displayed below in conjunction with (B.2):

$$\sqrt{\frac{\delta}{\Delta(\delta)}} = \frac{1}{\mathbb{E}[\bar{\Sigma}]}\mathbb{E}\Big[\Big(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1\Big)\mathcal{J}^*(\bar{Y})\Big]\mathbb{E}\Big[\frac{\bar{\Sigma}^2}{\gamma^\circ - \mathbb{E}[\mathcal{J}^*(\bar{Y})]\bar{\Sigma}}\Big],$$

$$1 = \frac{1}{\delta}\mathbb{E}[\mathcal{J}^*(\bar{Y})^2]\mathbb{E}\Big[\Big(\frac{\bar{\Sigma}}{\gamma^\circ - \mathbb{E}[\mathcal{J}^*(\bar{Y})]\bar{\Sigma}}\Big)^2\Big], \tag{B.9}$$

where

$$1 = \frac{1}{\delta} \mathbb{E}\left[\frac{\overline{\Sigma}}{\gamma^\circ - \mathbb{E}[\mathcal{J}^*(\overline{Y})]\overline{\Sigma}} \right]. \tag{B.10}$$

Note that the first identity in (B.9) implies (B.1) since $\delta > \Delta(\delta)$ by (B.5).

Let us verify the validity of (B.9). By the construction of $\mathcal{J}^*$ (see (B.8)),

$$\begin{aligned} \mathbb{E}\left[\mathcal{J}^*(\overline{Y})\right] &= \int_{\mathrm{supp}(\overline{Y})} m_0(y)\mathcal{J}^*(y)\,\mathrm{d}y \\ &= \sqrt{\frac{\Delta(\delta)}{\delta}} \int_{\mathrm{supp}(\overline{Y})} m_2(y) - m_0(y)\,\mathrm{d}y = 0, \end{aligned} \tag{B.11}$$

where the last equality follows from (B.7). Using this in (B.10), we can solve γ° explicitly:

$$\gamma^\circ = \frac{\mathbb{E}[\overline{\Sigma}]}{\delta}. \tag{B.12}$$

Consequently, the first two identities of (B.9) can be simplified as follows. First look at the first identity of (B.9). The right-hand side equals

$$\begin{aligned} &\frac{1}{\mathbb{E}[\overline{\Sigma}]} \mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\overline{\Sigma}]}\overline{G}^2 - 1\right)\mathcal{J}^*(\overline{Y})\right] \mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma^\circ - \mathbb{E}[\mathcal{J}^*(\overline{Y})]\overline{\Sigma}}\right] \\ &= \frac{\delta\mathbb{E}[\overline{\Sigma}^2]}{\mathbb{E}[\overline{\Sigma}]^2} \mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\overline{\Sigma}]}\overline{G}^2 - 1\right)\mathcal{J}^*(\overline{Y})\right] \tag{B.13} \\ &= \frac{\delta\mathbb{E}[\overline{\Sigma}^2]}{\mathbb{E}[\overline{\Sigma}]^2} \int_{\mathrm{supp}(\overline{Y})} \left(m_2(y) - m_0(y)\right)\mathcal{J}^*(y)\,\mathrm{d}y \\ &= \sqrt{\Delta(\delta)\delta} \cdot \frac{\mathbb{E}[\overline{\Sigma}^2]}{\mathbb{E}[\overline{\Sigma}]^2} \int_{\mathrm{supp}(\overline{Y})} \frac{(m_2(y) - m_0(y))^2}{m_0(y)}\,\mathrm{d}y. \tag{B.14} \end{aligned}$$

Equation (B.13) is by (B.11) and (B.12). Equation (B.14) is by (B.8). Therefore, the first identity of (B.9) is equivalent to

$$\Delta(\delta) = \frac{\mathbb{E}[\overline{\Sigma}]^2}{\mathbb{E}[\overline{\Sigma}^2]} \left(\int_{\mathrm{supp}(\overline{Y})} \frac{(m_2(y) - m_0(y))^2}{m_0(y)}\,\mathrm{d}y\right)^{-1}.$$

The right-hand side is the same as that of (B.5), hence the first identity of (B.9) indeed holds by the definition of $\Delta(\delta)$.

Next, we move to the second identity of (B.9). Using (B.11) and (B.12) again, the right-hand side equals

$$\begin{aligned} &\frac{1}{\delta} \mathbb{E}[\mathcal{J}^*(\overline{Y})^2] \mathbb{E}\left[\left(\frac{\overline{\Sigma}}{\gamma^\circ - \mathbb{E}[\mathcal{J}^*(\overline{Y})]\overline{\Sigma}}\right)^2\right] \\ &= \frac{1}{\delta} \mathbb{E}[\mathcal{J}^*(\overline{Y})^2] \frac{\mathbb{E}[\overline{\Sigma}^2]}{(\gamma^\circ)^2} = \delta \frac{\mathbb{E}[\overline{\Sigma}^2]}{\mathbb{E}[\overline{\Sigma}]^2} \mathbb{E}[\mathcal{J}^*(\overline{Y})^2] \end{aligned}$$

$$= \Delta(\delta) \frac{\mathbb{E}[\bar{\Sigma}^2]}{\mathbb{E}[\bar{\Sigma}]^2} \mathbb{E}\left[\left(\frac{m_2(\bar{Y})}{m_0(\bar{Y})} - 1\right)^2\right]$$

$$= \Delta(\delta) \frac{\mathbb{E}[\bar{\Sigma}^2]}{\mathbb{E}[\bar{\Sigma}]^2} \int_{\text{supp}(\bar{Y})} m_0(y)\left(\frac{m_2(y)}{m_0(y)} - 1\right)^2 \mathrm{d}y = 1,$$

which verifies the second identity of (B.9). The second line uses the definition of $\mathcal{J}^*$ in (B.8) and the last equality is by the definition of $\Delta(\delta)$ (see the right-hand side of (B.5)).

To complete the proof, it remains to verify that $\mathcal{T}^*$ satisfies assumption (A5). Recalling (B.6) and (B.8), we have

$$\mathcal{T}^*(y) = \frac{\mathcal{J}^*(y)}{1 + \mathcal{J}^*(y)} = \frac{\sqrt{\frac{\Delta(\delta)}{\delta}}\left(\frac{m_2(y)}{m_0(y)} - 1\right)}{1 + \sqrt{\frac{\Delta(\delta)}{\delta}}\left(\frac{m_2(y)}{m_0(y)} - 1\right)}$$

$$= 1 - \frac{1}{\sqrt{\frac{\Delta(\delta)}{\delta}} \frac{m_2(y)}{m_0(y)} + 1 - \sqrt{\frac{\Delta(\delta)}{\delta}}}.$$

By definitions, both m_2 and m_0 are non-negative functions. Therefore,

$$\inf_{y \in \text{supp}(\bar{Y})} \mathcal{T}^*(y) \geq 1 - \frac{1}{1 - \sqrt{\frac{\Delta(\delta)}{\delta}}} > -\infty, \tag{B.15}$$

where the last inequality holds since $\delta > \Delta(\delta)$ by the assumption in (B.5). Also, it trivially holds that

$$\sup_{y \in \text{supp}(\bar{Y})} \mathcal{T}^*(y) \leq 1 < \infty. \tag{B.16}$$

It is easy to see that $\mathcal{T}^*(y) > 0$ if and only if $m_2(y) > m_0(y)$. We first claim that m_2 and m_0 are not identically equal. Otherwise, $\Delta(\delta)$ (i.e., the right-hand side of (B.5)) is infinity and δ satisfying (B.5) is also infinity, violating assumption (A4). Moreover, by (B.7),

$$\int_{\text{supp}(\bar{Y})} m_2(y) - m_0(y) \, \mathrm{d}y = 0.$$

It follows from the mean value theorem for definite integrals that there exists $y \in \text{supp}(\bar{Y})$ such that $m_2(y) > m_0(y)$, which implies

$$\sup_{y \in \text{supp}(\bar{Y})} \mathcal{T}^*(y) > 0. \tag{B.17}$$

Since $\mathcal{T}^*$ is assumed to be pseudo-Lipschitz of finite order, putting (B.15) to (B.17) together verifies assumption (A5).

Note that, by the arguments in Section C, $\mathcal{T}^*$ does not need to satisfy assumption (A7) to have positive limiting overlap. In fact, if (3.13) holds and $\mathcal{T}^*$ does

not have a point mass at the boundaries of its support (otherwise assumption (A7) automatically holds), we can create such point masses via a perturbation. Now, the perturbed function satisfies assumption (A7) and it has positive limiting overlap for all sufficiently small perturbations. Then, an application of the Davis–Kahan theorem shows that we can set the perturbation to 0, and obtain the desired result for $\mathcal{T}^*$. This concludes the proof.

C. Removing assumptions (A6) and (A7)

We show that the conclusions of Theorem 3.1 remain valid even if Σ and/or $\mathcal{T}$ fail to satisfy assumption (A6) and/or (A7). To do so, we create $\widetilde{\Sigma}, \widetilde{\mathcal{T}}$ that closely approximate $\Sigma, \mathcal{T}$ and satisfy assumptions (A6) and (A7). Theorem 3.1 then applies to $\widetilde{\Sigma}, \widetilde{\mathcal{T}}$. We then show using a perturbation analysis that the same characterizations also hold for $\Sigma, \mathcal{T}$ once the perturbation is sent to zero. The detailed proof is presented below where we assume that both assumptions (A6) and (A7) are violated. The proof when only one of them holds is analogous and is omitted.

We first construct $\widetilde{\Sigma}$. Note that if

$$\mathbb{P}\big(\bar{\Sigma} = \inf \operatorname{supp}(\bar{\Sigma})\big) > 0, \quad \mathbb{P}\big(\bar{\Sigma} = \sup \operatorname{supp}(\bar{\Sigma})\big) > 0, \tag{C.1}$$

then assumption (A6) is automatically satisfied and one can take $\widetilde{\Sigma} = \Sigma$. In what follows, we assume that both probabilities in (C.1) are zero. (Again, the case where exactly one of the probabilities is zero can be handled verbatim and the details are omitted.) Write the eigendecomposition of Σ as $\Sigma = \sum_{i=1}^{d} \lambda_i(\Sigma) v_i(\Sigma) v_i(\Sigma)^\top$. By the convergence of the empirical spectral distribution of Σ (see assumption (A2)), we have that for any sufficiently small $\varsigma > 0$, there exists $\xi > 0$ (depending on ς) such that for every sufficiently large d,

$$\frac{1}{d}\big|\{i \in \{1, \ldots, d\} : \lambda_i(\Sigma) \geq \big(\sqrt{\lambda_1(\Sigma)} - \xi\big)^2\}\big| \in [\varsigma/2, \varsigma],$$

$$\frac{1}{d}\big|\{i \in \{1, \ldots, d\} : \lambda_i(\Sigma) \leq \big(\sqrt{\lambda_d(\Sigma)} + \xi\big)^2\}\big| \in [\varsigma/2, \varsigma].$$

Let $\widetilde{\Sigma} \in \mathbb{R}^{d \times d}$ be the matrix obtained by truncating the spectrum of Σ:

$$\widetilde{\Sigma} = \sum_{i=1}^{d} \lambda_i(\widetilde{\Sigma}) v_i(\Sigma) v_i(\Sigma)^\top,$$

where

$$\lambda_i(\widetilde{\Sigma}) = \begin{cases} \big(\sqrt{\lambda_1(\Sigma)} - \xi\big)^2, & \lambda_i(\Sigma) \geq \big(\sqrt{\lambda_1(\Sigma)} - \xi\big)^2, \\ \big(\sqrt{\lambda_d(\Sigma)} + \xi\big)^2, & \lambda_i(\Sigma) \leq \big(\sqrt{\lambda_d(\Sigma)} + \xi\big)^2, \\ \lambda_i(\Sigma), & \text{otherwise}. \end{cases}$$

It is easy to check that $\widetilde{\Sigma}$ still satisfies assumption (A2) if Σ does. Moreover, upon truncation, the limiting spectral distribution of $\widetilde{\Sigma}$ has positive mass on both the left and right edges, and hence obviously satisfies assumption (A6).

Let us then construct $\widetilde{\mathcal{T}}$. Clearly, if

$$\mathbb{P}\big(\mathcal{T}(\bar{Y}) = \sup \operatorname{supp}(\mathcal{T}(\bar{Y}))\big) > 0, \tag{C.2}$$

then (2.6) is satisfied. We therefore assume that the above equation holds with equality. In this case, we truncate $\mathcal{T}$ slightly below its supremum to create $\widetilde{\mathcal{T}}$ which satisfies (2.6). Specifically, for any $\varsigma > 0$, there exists $\xi > 0$ (depending on ς) such that

$$\mathbb{P}\big(\mathcal{T}(\bar{Y}) \in \big[\sup \operatorname{supp}(\mathcal{T}(\bar{Y})) - \xi, \sup \operatorname{supp}(\mathcal{T}(\bar{Y}))\big]\big) \in [\varsigma/2, \varsigma].$$

Define $\widetilde{\mathcal{T}}$ as

$$\widetilde{\mathcal{T}}(y) := \min\{\mathcal{T}(y), \sup \operatorname{supp}(\mathcal{T}(\bar{Y})) - \xi\}.$$

Note that $\widetilde{\mathcal{T}}$ depends on ς. Also, it satisfies (C.2) and therefore (2.6). It is easy to see that assumption (A5) will not be violated after the truncation.

Now the conclusions of Theorem 3.1 hold for $\widetilde{\Sigma}, \widetilde{\mathcal{T}}$. In particular, $\widetilde{a}^*, \widetilde{a}^\circ$ can be defined using (3.4) and (3.6) but with $\widetilde{\mathcal{T}}$ and the limiting spectral distribution of $\widetilde{\Sigma}$. It then suffices to show that as long as $\widetilde{a}^* > \widetilde{a}^\circ$, the difference between the spectral statistics under $\Sigma, \mathcal{T}$ and those under $\widetilde{\Sigma}, \widetilde{\mathcal{T}}$ is vanishing as $\varsigma \to 0$. Let

$$D := \Sigma^{1/2} \widetilde{X}^\top T \widetilde{X} \Sigma^{1/2}, \quad \widetilde{D} := \widetilde{\Sigma}^{1/2} \widetilde{X}^\top \widetilde{T} \widetilde{X} \widetilde{\Sigma}^{1/2},$$

where

$$T := \operatorname{diag}(\mathcal{T}(y)), \quad \widetilde{T} := \operatorname{diag}(\widetilde{\mathcal{T}}(y)).$$

Then

$$\begin{aligned}
\|D - \widetilde{D}\|_2 &= \|\Sigma^{1/2} \widetilde{X}^\top T \widetilde{X} \Sigma^{1/2} - \widetilde{\Sigma}^{1/2} \widetilde{X}^\top \widetilde{T} \widetilde{X} \widetilde{\Sigma}^{1/2}\|_2 \\
&\leq \|\Sigma^{1/2} \widetilde{X}^\top T \widetilde{X} \Sigma^{1/2} - \widetilde{\Sigma}^{1/2} \widetilde{X}^\top T \widetilde{X} \Sigma^{1/2}\|_2 \\
&\quad + \|\widetilde{\Sigma}^{1/2} \widetilde{X}^\top T \widetilde{X} \Sigma^{1/2} - \widetilde{\Sigma}^{1/2} \widetilde{X}^\top \widetilde{T} \widetilde{X} \Sigma^{1/2}\|_2 \\
&\quad + \|\widetilde{\Sigma}^{1/2} \widetilde{X}^\top \widetilde{T} \widetilde{X} \Sigma^{1/2} - \widetilde{\Sigma}^{1/2} \widetilde{X}^\top \widetilde{T} \widetilde{X} \widetilde{\Sigma}^{1/2}\|_2 \\
&\leq \|\Sigma^{1/2} - \widetilde{\Sigma}^{1/2}\|_2 \|\widetilde{X}\|_2^2 \|T\|_2 \|\Sigma^{1/2}\|_2 \\
&\quad + \|\widetilde{\Sigma}^{1/2}\|_2 \|\widetilde{X}\|_2^2 \|T - \widetilde{T}\|_2 \|\Sigma^{1/2}\|_2 \\
&\quad + \|\widetilde{\Sigma}^{1/2}\|_2 \|\widetilde{X}\|_2^2 \|\widetilde{T}\|_2 \|\Sigma^{1/2} - \widetilde{\Sigma}^{1/2}\|_2 \\
&\leq 2\|\Sigma^{1/2} - \widetilde{\Sigma}^{1/2}\|_2 \|\widetilde{X}\|_2^2 \|T\|_2 \|\Sigma^{1/2}\|_2 \\
&\quad + \|\Sigma^{1/2}\|_2^2 \|\widetilde{X}\|_2^2 \|T - \widetilde{T}\|_2
\end{aligned}$$

$$\leq 2\xi\left(1 + \frac{1}{\sqrt{\delta}} + 0.01\right)^2 \left(\sup \operatorname{supp}(\mathcal{T}(\bar{Y})) + 0.01\right)\left(\operatorname{supp}(\bar{\Sigma}) + 0.01\right)$$

$$+ \left(\operatorname{supp}(\bar{\Sigma}) + 0.01\right)\left(1 + \frac{1}{\sqrt{\delta}} + 0.01\right)^2 \xi$$

$$\leq c_1 \xi, \tag{C.3}$$

where the bound on the penultimate line holds almost surely for every sufficiently large d, and $c_1 > 0$ in the last line is a constant independent of d. The $+0.01$ terms are to exclude deviations for small d. Furthermore, if $\tilde{a}^* > \tilde{a}^\circ$, Theorem 3.1 guarantees that there exists a constant $c_2 > 0$ such that for every sufficiently large d, with probability 1,

$$\lambda_1(\tilde{D}) - \lambda_2(\tilde{D}) \geq c_2. \tag{C.4}$$

Using (C.3) and (C.4) in the Davis–Kahan theorem (Proposition G.4), we obtain

$$\min\{\|v_1(D) - v_1(\tilde{D})\|_2, \|v_1(D) + v_1(\tilde{D})\|_2\} \leq \frac{4\|D - \tilde{D}\|_2}{\lambda_1(D) - \lambda_2(D)} \leq 4c_1\xi/c_2,$$

which implies

$$\left|\left|\left\langle v_1(D), \frac{\beta^*}{\sqrt{d}}\right\rangle\right| - \left|\left\langle v_1(\tilde{D}), \frac{\beta^*}{\sqrt{d}}\right\rangle\right|\right| = \min_{\sigma \in \{-1,1\}}\left|\left\langle v_1(D) - \sigma v_1(\tilde{D}), \frac{\beta^*}{\sqrt{d}}\right\rangle\right|$$

$$\leq \min_{\sigma \in \{-1,1\}} \|v_1(D) - v_1(\tilde{D})\|_2$$

$$\leq 4c_1\xi/c_2. \tag{C.5}$$

By Theorem 3.1, the condition $\tilde{a}^* > \tilde{a}^\circ$ also implies that the overlap between $v_1(\tilde{D})$ and β^* converges in probability to $\eta > 0$. Since $\varsigma > 0$ (and therefore ξ) can be made arbitrarily small, (C.5) then allows us to conclude that the overlap between $v_1(D)$ and β^* also converges to η. This proves (3.12) for D.

Using (C.3) and Weyl's inequality, we have for any $i \in \{1, \ldots, d\}$,

$$|\lambda_i(D) - \lambda_i(\tilde{D})| \leq \|D - \tilde{D}\|_2 \leq c_1\xi,$$

which in particular establishes (3.11) for D. This completes the proof.

D. Properties of auxiliary functions and parameters

D.1. Existence and uniqueness of a^*

Recall the functions $\varphi, \psi\colon (\sup \operatorname{supp}(\mathcal{T}(\bar{Y})), \infty) \to \mathbb{R}$ defined in (3.2).

Proposition D.1 (Existence of a^*). *Let assumption* (A7) *hold. Then, the equation* $\varphi(a^*) = \zeta(a^*)$ *has at least one solution in* $(\sup \operatorname{supp}(\mathcal{T}(\bar{Y})), \infty)$.

Proof. Recall that both φ and ζ are defined on $(\sup \text{supp}(\mathcal{T}(\bar{Y})), \infty)$. It is not hard to see from (3.3) that γ is a continuous function. Therefore, φ, ψ, ζ are also continuous. We will show

$$\lim_{a \searrow \sup \text{supp}(\mathcal{T}(\bar{Y}))} \varphi(a) > \lim_{a \searrow \sup \text{supp}(\mathcal{T}(\bar{Y}))} \zeta(a), \quad \lim_{a \nearrow \infty} \varphi(a) < \lim_{a \nearrow \infty} \zeta(a). \quad \text{(D.1)}$$

Then by the intermediate value theorem, this immediately implies the result.

We will explicitly evaluate the four limits. To this end, let us first study the limiting values of $\gamma(a)$ defined through (3.3).

Limiting values of γ. By inspecting the defining equation, it is clear that

$$\lim_{a \to \infty} \frac{1}{\delta} \mathbb{E}\left[\frac{\bar{\Sigma}}{\gamma - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\Sigma}} \right] = \frac{\mathbb{E}[\bar{\Sigma}]}{\delta \gamma},$$

and hence

$$\lim_{a \to \infty} \gamma(a) = \frac{\mathbb{E}[\bar{\Sigma}]}{\delta}, \quad \text{(D.2)}$$

which is positive and finite. We also claim that

$$\lim_{a \searrow \sup \text{supp}(\mathcal{T}(\bar{Y}))} \gamma(a) = \infty. \quad \text{(D.3)}$$

Otherwise, for any finite γ, by (d) in (2.6),

$$\lim_{a \searrow \sup \text{supp}(\mathcal{T}(\bar{Y}))} \frac{1}{\delta} \mathbb{E}\left[\frac{\bar{\Sigma}}{\gamma - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\Sigma}} \right] = 0,$$

which violates (3.3). The possibility of

$$\lim_{a \searrow \sup \text{supp}(\mathcal{T}(\bar{Y}))} \gamma(a) = -\infty$$

can be similarly excluded.

Limiting values of φ. We claim that

$$\lim_{a \searrow \sup \text{supp}(\mathcal{T}(\bar{Y}))} \varphi(a) = \infty, \quad \lim_{a \to \infty} \varphi(a) = \delta \mathbb{E}[\bar{G}^2 \mathcal{T}(\bar{Y})] \frac{\mathbb{E}[\bar{\Sigma}^2]}{\mathbb{E}[\bar{\Sigma}]^2} < \infty. \quad \text{(D.4)}$$

The limit towards the right boundary of the domain is easy to verify:

$$\begin{aligned}
\lim_{a \to \infty} \varphi(a) &= \lim_{a \to \infty} \frac{1}{\mathbb{E}[\bar{\Sigma}]} \mathbb{E}\left[\bar{G}^2 \frac{\mathcal{T}(\bar{Y})}{1 - \mathcal{T}(\bar{Y})/a} \right] \mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\Sigma}} \right] \\
&= \frac{1}{\mathbb{E}[\bar{\Sigma}]} \mathbb{E}[\bar{G}^2 \mathcal{T}(\bar{Y})] \mathbb{E}\left[\frac{\bar{\Sigma}^2}{\mathbb{E}[\bar{\Sigma}]/\delta} \right] \\
&= \delta \mathbb{E}[\bar{G}^2 \mathcal{T}(\bar{Y})] \frac{\mathbb{E}[\bar{\Sigma}^2]}{\mathbb{E}[\bar{\Sigma}]^2},
\end{aligned}$$

where we use (D.3) in the second equality. To show the first equality in (D.4), let us start by observing that for any $a > \sup \operatorname{supp}(\mathcal{T}(\bar{Y}))$,

$$
\begin{aligned}
0 &< \mathbb{E}\left[\frac{1}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\bar{\Sigma}}}\right] \\
&\le \frac{1}{\inf \operatorname{supp}(\bar{\Sigma})} \mathbb{E}\left[\frac{\bar{\Sigma}}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\bar{\Sigma}}}\right] = \frac{\delta}{\inf \operatorname{supp}(\bar{\Sigma})}.
\end{aligned}
\tag{D.5}
$$

The second inequality is valid since $\inf \operatorname{supp}(\bar{\Sigma}) > 0$ by assumption (A2), and hence

$$
\frac{\bar{\Sigma}}{\inf \operatorname{supp}(\bar{\Sigma})} \ge 1
$$

almost surely. The last equality is by the definition of $\gamma(\cdot)$ (see (3.3)). On the other hand, a simple application of the Cauchy–Schwarz inequality yields:

$$
\begin{aligned}
\delta^2 &= \mathbb{E}\left[\frac{\bar{\Sigma}}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\bar{\Sigma}}}\right]^2 \\
&\le \mathbb{E}\left[\frac{\bar{\Sigma}}{(\gamma(a) - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\bar{\Sigma}})^{1/2}} \cdot \frac{1}{(\gamma(a) - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\bar{\Sigma}})^{1/2}}\right]^2 \\
&\le \mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\bar{\Sigma}}}\right] \mathbb{E}\left[\frac{1}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\bar{\Sigma}}}\right].
\end{aligned}
$$

Rearranging and using (D.5) gives:

$$
\mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\bar{\Sigma}}}\right] \ge \frac{\delta^2}{\mathbb{E}[1/(\gamma(a) - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\bar{\Sigma}})]} \ge \delta \cdot \inf \operatorname{supp}(\bar{\Sigma}),
$$

the right-hand side of which is a strictly positive lower bound independent of a. From here, we conclude

$$
\begin{aligned}
\lim_{a \searrow \sup \operatorname{supp}(\mathcal{T}(\bar{Y}))} \varphi(a) &= \lim_{a \searrow \sup \operatorname{supp}(\mathcal{T}(\bar{Y}))} \frac{a}{\mathbb{E}[\bar{\Sigma}]} \mathbb{E}[\bar{G}^2 \mathcal{F}_a(\bar{Y})] \mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\bar{\Sigma}}}\right] \\
&= \infty,
\end{aligned}
$$

since the middle term converges to ∞ by (e) in (2.6) and the remaining terms are lower bounded by some positive constant as $a \searrow \sup \operatorname{supp}(\mathcal{T}(\bar{Y}))$.

Limiting values of ζ. By definition,

$$
\lim_{a \searrow \sup \operatorname{supp}(\mathcal{T}(\bar{Y}))} \zeta(a) = \zeta(a^\circ) = \psi(a^\circ) < \infty.
\tag{D.6}
$$

Using (D.2), we obtain

$$
\lim_{a \to \infty} \zeta(a) = \lim_{a \to \infty} \psi(a) = \lim_{a \to \infty} a\gamma(a) = \infty.
\tag{D.7}
$$

Finally, combining (D.4), (D.6), and (D.7) gives (D.1), which completes the proof of the proposition. ∎

Proposition D.2 (Monotonicity of φ). *Let assumption* (A5) *hold. Suppose*

$$\inf_{y \in \mathrm{supp}(\bar{Y})} \mathcal{T}(y) \geq 0. \tag{D.8}$$

Then, the function φ is strictly decreasing.

Proof. We show that φ is strictly decreasing by proving $\varphi' < 0$. Let us start by computing φ'. Recall

$$\mathbb{E}[\bar{\Sigma}]\varphi(a) = \mathbb{E}\big[\bar{G}^2 a \mathcal{F}_a(\bar{Y})\big]\mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\Sigma}}\right].$$

Using the chain rule, we obtain

$$\mathbb{E}[\bar{\Sigma}]\varphi'(a) = -\mathbb{E}\big[\bar{G}^2 \mathcal{F}_a(\bar{Y})^2\big]\mathbb{E}\left[\frac{\bar{\Sigma}^2}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\Sigma}}\right]$$
$$- \mathbb{E}\big[\bar{G}^2 a \mathcal{F}_a(\bar{Y})\big]\mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma(a) - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\Sigma})^2}\left(\gamma'(a) + \mathbb{E}\left[\frac{\mathcal{T}(\bar{Y})}{(a - \mathcal{T}(\bar{Y}))^2}\right]\bar{\Sigma}\right)\right]. \tag{D.9}$$

The derivative of γ can be accessed via the implicit function theorem. Let

$$H(a, \gamma) = \frac{1}{\delta}\mathbb{E}\left[\frac{\bar{\Sigma}}{\gamma - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\Sigma}}\right] - 1.$$

Recalling (3.3), we see that $\gamma(a)$ is the solution γ to $H(a, \gamma) = 0$. We have

$$\frac{\partial}{\partial a}H(a, \gamma) = \frac{1}{\delta}\mathbb{E}\left[\frac{-\bar{\Sigma}}{(\gamma - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\Sigma})^2} \cdot (-\bar{\Sigma}) \cdot \mathbb{E}\left[\frac{-\mathcal{T}(\bar{Y})}{(a - \mathcal{T}(\bar{Y}))^2}\right]\right]$$
$$= -\frac{1}{\delta}\mathbb{E}\left[\frac{\mathcal{T}(\bar{Y})}{(a - \mathcal{T}(\bar{Y}))^2}\right]\mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\Sigma})^2}\right],$$

and

$$\frac{\partial}{\partial \gamma}H(a, \gamma) = -\frac{1}{\delta}\mathbb{E}\left[\frac{\bar{\Sigma}}{(\gamma - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\Sigma})^2}\right].$$

By the implicit function theorem,

$$\frac{\mathrm{d}}{\mathrm{d}a}\gamma(a) = -\frac{\frac{\partial}{\partial a}H(a, \gamma(a))}{\frac{\partial}{\partial \gamma}H(a, \gamma(a))}$$
$$= -\frac{\mathbb{E}[\mathcal{T}(\bar{Y})/(a - \mathcal{T}(\bar{Y}))^2]\mathbb{E}[\bar{\Sigma}^2/(\gamma(a) - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\Sigma})^2]}{\mathbb{E}[\bar{\Sigma}/(\gamma(a) - \mathbb{E}[\mathcal{F}_a(\bar{Y})]\bar{\Sigma})^2]}. \tag{D.10}$$

Using this, we simplify the second term of (D.9):

$$
-\mathbb{E}\big[\bar{G}^2 a \mathscr{F}_a(\bar{Y})\big]\mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma(a)-\mathbb{E}[\mathscr{F}_a(\bar{Y})]\bar{\Sigma})^2}\left(\gamma'(a)+\mathbb{E}\left[\frac{\mathcal{T}(\bar{Y})}{(a-\mathcal{T}(\bar{Y}))^2}\right]\bar{\Sigma}\right)\right]
$$

$$
= -\mathbb{E}\big[\bar{G}^2 a \mathscr{F}_a(\bar{Y})\big]\mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma(a)-\mathbb{E}[\mathscr{F}_a(\bar{Y})]\bar{\Sigma})^2}\right]\gamma'(a)
$$

$$
- \mathbb{E}\big[\bar{G}^2 a \mathscr{F}_a(\bar{Y})\big]\mathbb{E}\left[\frac{\mathcal{T}(\bar{Y})}{(a-\mathcal{T}(\bar{Y}))^2}\right]\mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma(a)-\mathbb{E}[\mathscr{F}_a(\bar{Y})]\bar{\Sigma})^2}\right]
$$

$$
= \mathbb{E}\big[\bar{G}^2 a \mathscr{F}_a(\bar{Y})\big]\mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma(a)-\mathbb{E}[\mathscr{F}_a(\bar{Y})]\bar{\Sigma})^2}\right]^2\frac{\mathbb{E}[\mathcal{T}(\bar{Y})/(a-\mathcal{T}(\bar{Y}))^2]}{\mathbb{E}[\bar{\Sigma}/(\gamma(a)-\mathbb{E}[\mathscr{F}_a(\bar{Y})]\bar{\Sigma})^2]}
$$

$$
- \mathbb{E}\big[\bar{G}^2 a \mathscr{F}_a(\bar{Y})\big]\mathbb{E}\left[\frac{\mathcal{T}(\bar{Y})}{(a-\mathcal{T}(\bar{Y}))^2}\right]\mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma(a)-\mathbb{E}[\mathscr{F}_a(\bar{Y})]\bar{\Sigma})^2}\right]. \quad \text{(D.11)}
$$

Let us argue that the right-hand side is negative. First note that since

(i) $a > \sup \operatorname{supp}(\mathcal{T}(\bar{Y})) > 0$,

(ii) $\inf \operatorname{supp}(\mathcal{T}(\bar{Y})) \geq 0$ by (D.8), and

(iii) $\mathcal{T}(\bar{Y})$ is not almost surely zero by assumption (A5),

the common factors are positive:

$$
\mathbb{E}\big[\bar{G}^2 a \mathscr{F}_a(\bar{Y})\big]\mathbb{E}\left[\frac{\mathcal{T}(\bar{Y})}{(a-\mathcal{T}(\bar{Y}))^2}\right] > 0. \quad \text{(D.12)}
$$

Then we apply the Cauchy–Schwarz inequality to obtain:

$$
\mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma(a)-\mathbb{E}[\mathscr{F}_a(\bar{Y})]\bar{\Sigma})^2}\right]^2
$$

$$
= \mathbb{E}\left[\frac{\bar{\Sigma}^{1/2}}{\gamma(a)-\mathbb{E}[\mathscr{F}_a(\bar{Y})]\bar{\Sigma}} \cdot \frac{\bar{\Sigma}^{3/2}}{\gamma(a)-\mathbb{E}[\mathscr{F}_a(\bar{Y})]\bar{\Sigma}}\right]^2 \quad \text{(D.13)}
$$

$$
\leq \mathbb{E}\left[\frac{\bar{\Sigma}}{(\gamma(a)-\mathbb{E}[\mathscr{F}_a(\bar{Y})]\bar{\Sigma})^2}\right]\mathbb{E}\left[\frac{\bar{\Sigma}^3}{(\gamma(a)-\mathbb{E}[\mathscr{F}_a(\bar{Y})]\bar{\Sigma})^2}\right]. \quad \text{(D.14)}
$$

Equation (D.13) is valid since $\bar{\Sigma}$ is positive and $\gamma(a) > s(a)$. The inequalities (D.12) and (D.14) jointly imply that the right-hand side of (D.11), i.e., the second term of (D.9), is non-positive, as claimed. Moreover, the first term of (D.9) is strictly negative. We therefore conclude that $\varphi'(a) < 0$ for any $a > \sup \operatorname{supp}(\mathcal{T}(\bar{Y}))$. ∎

Remark D.1 (Monotonicity of φ). The monotonicity property of φ relies on the non-negativity of $\mathcal{T}$ in (D.8). We believe that this assumption can be relaxed. In fact, numerical evidence suggests that φ is monotone: we report in Figure 6 that in the
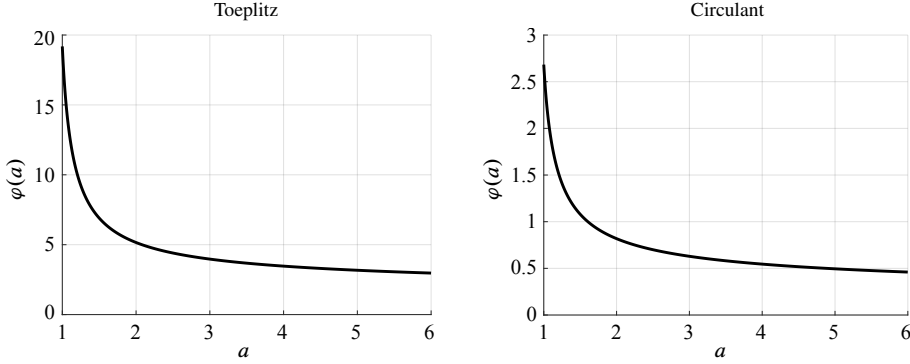
Figure 6. Plots of the function φ defined in (3.2) with parameters specified in Remark D.1.

setting of noiseless phase retrieval $q(g, \varepsilon) = |g|$ with optimal preprocessing function

$$\mathcal{T}(y) = \max\left\{1 - \frac{1}{\delta y^2}, -10\right\}$$

(where $\delta = 0.1$), the function φ is strictly decreasing and convex in $(1, \infty)$ (note that sup supp($\mathcal{T}(\bar{Y})$) = 1) when $\bar{\Sigma}$ is Toeplitz with $\rho = 0.9$ or circulant with $c_0 = 1$, $c_1 = 0.1$, $\ell = 17$. Note that the function $\mathcal{T}$ here is not everywhere non-negative.

Proposition D.3 (Uniqueness of a^*). *Let assumption* (A5) *hold. Suppose φ is strictly decreasing. Then, $\varphi(a^*) = \zeta(a^*)$ has a unique solution in* (sup supp($\mathcal{T}(\bar{Y})$), ∞).

Proof. The uniqueness of a^* follows from several properties that have been proved for φ and ζ. Recall the assumption that φ is strictly decreasing and that ζ is non-decreasing by Lemma E.1. Furthermore, from the proof of Proposition D.1 (in particular, (D.4), (D.6), and (D.7)), we know that in the interval (sup supp($\mathcal{T}(\bar{Y})$), ∞)), φ strictly decreases from ∞ to a finite constant, whereas ζ increases from a finite constant to ∞. By the intermediate value theorem, the solution to $\varphi(a^*) = \zeta(a^*)$ must exist and is unique. ∎

D.2. Equivalent definitions of $a°$, a^* and equivalent description of sup supp($\bar{\mu}_{\hat{D}}$)

Let $\mathcal{A} \subset \mathbb{R}^2$ be the domain on which the potential solutions to various self-consistent equations of interest are to be considered:

$$\mathcal{A} := \left\{(a, \gamma) : a > \text{sup supp}(\mathcal{T}(\bar{Y})), \gamma > s(a)\right\},$$

where $s(a)$ is defined in (3.1).

Proposition D.4 (Equivalent definitions of a°, a^*). *The following statements hold:*

- *In the domain $\mathcal{A}$, the unique solution (a°, γ°) to*

$$
1 = \frac{1}{\delta} \mathbb{E}\big[\mathscr{F}_{a^\circ}(\bar{Y})^2\big] \mathbb{E}\left[\left(\frac{\bar{\Sigma}}{\gamma^\circ - \mathbb{E}[\mathscr{F}_{a^\circ}(\bar{Y})]\bar{\Sigma}}\right)^2\right],
$$
$$
1 = \frac{1}{\delta} \mathbb{E}\left[\frac{\bar{\Sigma}}{\gamma^\circ - \mathbb{E}[\mathscr{F}_{a^\circ}(\bar{Y})]\bar{\Sigma}}\right]
\tag{D.15}
$$

is the same as the unique solution to the following equations:

$$
\psi'(a^\circ) = 0, \quad \gamma^\circ = \gamma(a^\circ).
\tag{D.16}
$$

- *Let (a^*, γ^*) be the solution in $\mathcal{A}$ to*

$$
\zeta(a^*) = \varphi(a^*), \quad \gamma^* = \gamma(a^*),
\tag{D.17}
$$

such that a^* is the largest among all solutions. If $a^* > a^\circ$, then (a^*, γ^*) is also a solution to (5.4).

Proof. We start by showing the equivalence between (D.15) and (D.16). We will argue that $\psi'(a) = 0$ if and only if (D.15) holds. The derivative of ψ' is

$$
\psi'(a) = \gamma(a) + a\gamma'(a)
$$
$$
= \gamma(a) - a \cdot \frac{\mathbb{E}[\mathscr{T}(\bar{Y})/(a - \mathscr{T}(\bar{Y}))^2]\mathbb{E}[\bar{\Sigma}^2/(\gamma(a) - \mathbb{E}[\mathscr{F}_a(\bar{Y})]\bar{\Sigma})^2]}{\mathbb{E}[\bar{\Sigma}/(\gamma(a) - \mathbb{E}[\mathscr{F}_a(\bar{Y})]\bar{\Sigma})^2]},
$$

where the formula for γ' has been derived in (D.10). Using the above expression and rearranging terms, we can write the equation $\psi'(a) = 0$ as

$$
\mathbb{E}\left[\frac{\gamma(a)\bar{\Sigma}}{(\gamma(a) - \mathbb{E}[\mathscr{F}_a(\bar{Y})]\bar{\Sigma})^2}\right] = \mathbb{E}\left[\frac{a\mathscr{T}(\bar{Y})}{(a - \mathscr{T}(\bar{Y}))^2}\right]
$$
$$
\cdot \mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma(a) - \mathbb{E}[\mathscr{F}_a(\bar{Y})]\bar{\Sigma})^2}\right].
\tag{D.18}
$$

We rewrite the first two terms in the above equation in the following way:

$$
\mathbb{E}\left[\frac{\gamma(a)\bar{\Sigma}}{(\gamma(a) - \mathbb{E}[\mathscr{F}_a(\bar{Y})]\bar{\Sigma})^2}\right] = \mathbb{E}\left[\frac{\bar{\Sigma}}{\gamma(a) - \mathbb{E}[\mathscr{F}_a(\bar{Y})]\bar{\Sigma}}\right]
$$
$$
+ \mathbb{E}\left[\frac{\bar{\Sigma}^2}{(\gamma(a) - \mathbb{E}[\mathscr{F}_a(\bar{Y})]\bar{\Sigma})^2}\right]\mathbb{E}\big[\mathscr{F}_a(\bar{Y})\big],
$$
$$
\mathbb{E}\left[\frac{a\mathscr{T}(\bar{Y})}{(a - \mathscr{T}(\bar{Y}))^2}\right] = \mathbb{E}\big[\mathscr{F}_a(\bar{Y})\big] + \mathbb{E}\big[\mathscr{F}_a(\bar{Y})^2\big].
$$

Using the right-hand sides above in place of the left-hand sides in (D.18), we see that the term $\mathbb{E}[\overline{\Sigma}^2/(\gamma(a) - \mathbb{E}[\mathcal{F}_a(\overline{Y})]\overline{\Sigma})^2]\mathbb{E}[\mathcal{F}_a(\overline{Y})]$ cancels on both sides and (D.18) becomes

$$\mathbb{E}\left[\frac{\overline{\Sigma}}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\overline{Y})]\overline{\Sigma}}\right] = \mathbb{E}[\mathcal{F}_a(\overline{Y})^2]\mathbb{E}\left[\frac{\overline{\Sigma}^2}{(\gamma(a) - \mathbb{E}[\mathcal{F}_a(\overline{Y})]\overline{\Sigma})^2}\right].$$

The left-hand side equals δ since $\gamma(a)$ satisfies (3.3). Therefore, the above equation matches (D.15).

Next, assuming that (D.17) holds, we verify (5.4). For any $a > a°$,

$$\zeta(a) = \psi(a),$$

hence (D.17) can be written as

$$\frac{1}{\mathbb{E}[\overline{\Sigma}]}\mathbb{E}[\overline{G}^2\mathcal{F}_a(\overline{Y})]\mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\overline{Y})]\overline{\Sigma}}\right] = \gamma(a),$$

or equivalently,

$$\frac{1}{\mathbb{E}[\overline{\Sigma}]}\mathbb{E}\left[\frac{\delta}{\mathbb{E}[\overline{\Sigma}]}\overline{G}^2\mathcal{F}_a(\overline{Y})\right]\mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\overline{Y})]\overline{\Sigma}}\right] = \frac{\delta\gamma(a)}{\mathbb{E}[\overline{\Sigma}]}.$$

To show that the above equation is the same as (5.4), it suffices to verify

$$\frac{\delta\gamma(a)}{\mathbb{E}[\overline{\Sigma}]} = \frac{1}{\mathbb{E}[\overline{\Sigma}]}\mathbb{E}[\mathcal{F}_a(\overline{Y})]\mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\overline{Y})]\overline{\Sigma}}\right] + 1. \qquad \text{(D.19)}$$

We rewrite the first term on the right-hand side as

$$\frac{1}{\mathbb{E}[\overline{\Sigma}]}\mathbb{E}[\mathcal{F}_a(\overline{Y})]\mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\overline{Y})]\overline{\Sigma}}\right]$$

$$= \frac{1}{\mathbb{E}[\overline{\Sigma}]}\frac{1}{\mathbb{E}[\mathcal{F}_a(\overline{Y})]}$$
$$\cdot\left(\mathbb{E}\left[\frac{\mathbb{E}[\mathcal{F}_a(\overline{Y})]^2\overline{\Sigma}^2 - \gamma(a)\mathbb{E}[\mathcal{F}_a(\overline{Y})]\overline{\Sigma}}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\overline{Y})]\overline{\Sigma}}\right] + \mathbb{E}\left[\frac{\gamma(a)\mathbb{E}[\mathcal{F}_a(\overline{Y})]\overline{\Sigma}}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\overline{Y})]\overline{\Sigma}}\right]\right)$$

$$= \frac{1}{\mathbb{E}[\overline{\Sigma}]}\frac{1}{\mathbb{E}[\mathcal{F}_a(\overline{Y})]}$$
$$\cdot\left(-\mathbb{E}[\mathbb{E}[\mathcal{F}_a(\overline{Y})]\overline{\Sigma}] + \gamma(a)\mathbb{E}[\mathcal{F}_a(\overline{Y})]\mathbb{E}\left[\frac{\overline{\Sigma}}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\overline{Y})]\overline{\Sigma}}\right]\right)$$

$$= \frac{1}{\mathbb{E}[\overline{\Sigma}]}\left(\gamma(a)\mathbb{E}\left[\frac{\overline{\Sigma}}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\overline{Y})]\overline{\Sigma}}\right] - \mathbb{E}[\overline{\Sigma}]\right)$$

$$= \frac{\gamma(a)}{\mathbb{E}[\overline{\Sigma}]}\mathbb{E}\left[\frac{\overline{\Sigma}}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\overline{Y})]\overline{\Sigma}}\right] - 1.$$

Noting that $\gamma(a)$ satisfies (3.3), we further obtain

$$\frac{1}{\mathbb{E}[\overline{\Sigma}]}\mathbb{E}\big[\mathcal{F}_a(\overline{Y})\big]\mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma(a) - \mathbb{E}[\mathcal{F}_a(\overline{Y})]\overline{\Sigma}}\right] = \frac{\delta\gamma(a)}{\mathbb{E}[\overline{\Sigma}]} - 1.$$

This then implies (D.19), and hence (5.4). ∎

Finally, we derive an alternative form of (5.33) in terms of a°, γ° defined through a pair of self-consistent equations. The proof follows from verifying that $\psi'(a^\circ) = 0$ is algebraically equivalent to (D.15), as shown in Proposition D.4 above.

Lemma D.5. *The description of* $\sup \mathrm{supp}(\overline{\mu}_{\widehat{D}})$ *in Lemma 5.5 is equivalent to*

$$\sup \mathrm{supp}(\overline{\mu}_{\widehat{D}}) = a^\circ \gamma^\circ,$$

where $(a^\circ, \gamma^\circ) \in \mathcal{A}$ *solves* (D.15)*, and* a° *is the largest among all such solutions.*

D.3. Alternative formulations of $a^* > a^\circ$

The following proposition is a direct consequence of the monotonicity properties of ψ, φ (see Proposition D.2 and Lemma E.1).

Proposition D.6. *The following conditions are equivalent:*

(1) $a^* > a^\circ$.

(2) $\zeta(a^*) > \zeta(a^\circ)$.

(3) $\zeta'(a^*) > 0$, *or more explicitly*

$$1 > \frac{1}{\delta}\mathbb{E}\big[\mathcal{F}_{a^*}(\overline{Y})^2\big]\mathbb{E}\left[\frac{\overline{\Sigma}^2}{(\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma})^2}\right],$$

i.e., $1 > w_2$ *by recalling the definition of* w_2 *in* (3.10).

(4) *If the function* $\varphi \colon (\sup \mathrm{supp}(\mathcal{T}(\overline{Y})), \infty) \to \mathbb{R}$ *defined in* (3.2) *is strictly decreasing, the above conditions are further equivalent to* $\psi(a^\circ) < \varphi(a^\circ)$, *or more explicitly*

$$1 < \frac{1}{\mathbb{E}[\overline{\Sigma}]}\mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\overline{\Sigma}]}\overline{G}^2 - 1\right)\mathcal{F}_{a^\circ}(\overline{Y})\right]\mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma^\circ - \mathbb{E}[\mathcal{F}_{a^\circ}(\overline{Y})]\overline{\Sigma}}\right].$$

E. Proof of Lemma 5.5

Recall from (5.32) the definition of $\widehat{D} \in \mathbb{R}^{d \times d}$:

$$\widehat{D} = \Sigma^{1/2}\widehat{X}^\top T \widehat{X}\Sigma^{1/2}.$$

We already know that both $\lambda_1(\widehat{D})$ and $\lambda_3(\widehat{D})$ converge to the upper edge $\lambda^\circ = \sup \operatorname{supp} \bar\mu_{\widehat{D}}$ of the limiting spectrum (see Lemmas A.3 and A.4). The main goal of this section is to prove the characterization of the upper edge λ° in Lemma 5.5. We deduce Lemma 5.5 from the following lemma. We present the proofs of Lemmas 5.5 and E.1 at the end of this appendix.

Lemma E.1. *Let $a \in (\sup \operatorname{supp} \bar\mu_T, \infty)$. Then, the following hold:*

(1) *If $\psi(\tilde{a}) > \lambda^\circ$ for all $\tilde{a} \geq a$, then $\psi'(a) > 0$.*

(2) *If $\psi'(a) > 0$, then $\psi(a) \notin \operatorname{supp} \bar\mu_{\widehat{D}}$.*

We will see in Lemma E.3 that a° is indeed well defined. More precisely, ψ is an analytic function with at least one critical point, and $\psi'(a)$ converges to a positive number as $a \to \infty$.

E.1. Properties of ψ

Recall that $\psi \colon (\sup \operatorname{supp} \bar\mu_T, \infty) \to \mathbb{R}$ is defined by $\psi(a) = a\gamma(a)$. With a slight modification to the definition of $\gamma(a)$, we have the following result.

Lemma E.2. *The following statements hold:*

(1) *The sets $\mathcal{S}, \mathcal{S}' \subset \mathbb{R}$ defined by*

$$\mathcal{S} := \left\{a > \sup \operatorname{supp} \bar\mu_T : \mathbb{E}\big[\mathcal{F}_a(\bar Y)\big] = 0\right\},$$
$$\mathcal{S}' := \left\{a > \sup \operatorname{supp} \bar\mu_T : -\mathbb{E}\big[\mathcal{F}_a(\bar Y)\big] = \frac{1}{\delta}\right\}$$

are finite.

(2) *For each $a \in (\sup \operatorname{supp} \bar\mu_T, \infty) \setminus \mathcal{S}$, there exists a unique*

$$\omega \equiv \omega(a) \in \mathbb{R} \setminus \big(\inf \operatorname{supp} \bar\mu_\Sigma, \sup \operatorname{supp} \bar\mu_\Sigma\big)$$

such that

$$\delta \int_{\mathbb{R}} \frac{t}{t-a} \, d\bar\mu_T(t) = \int_{\mathbb{R}} \frac{s}{s-\omega} \, d\bar\mu_\Sigma(s). \tag{E.1}$$

(3) *The map $\omega \colon (\sup \operatorname{supp} \bar\mu_T, \infty) \setminus \mathcal{S} \to \mathbb{R}$ defined in item (2) extends meromorphically to an open set in $\mathbb{C}$ containing $(\sup \operatorname{supp} \bar\mu_T, \infty)$. The extension is analytic at each $a \in (\sup \operatorname{supp} \bar\mu_T, \infty) \setminus \mathcal{S}$, has a pole at each $a \in \mathcal{S}$ and a zero at each $a \in \mathcal{S}'$.*

(4) *The function $\psi \colon (\sup \operatorname{supp} \bar\mu_T, \infty) \to \mathbb{R}$ defined by $\psi(a) = a\gamma(a)$ satisfies*

$$\psi(a) = -\frac{a}{\delta} \int_{\mathbb{R}} \frac{s\omega(a)}{s - \omega(a)} \, d\bar\mu_\Sigma(s), \quad \forall a \in (\sup \operatorname{supp} \bar\mu_T, \infty) \setminus \mathcal{S}. \tag{E.2}$$

Furthermore, the function ψ extends analytically to an open set in $\mathbb{C}$ containing $(\sup \operatorname{supp} \bar{\mu}_T, \infty)$, and has zeros precisely at $\mathcal{S}'$.

Proof. Note that the function $a \mapsto -\mathbb{E}[\mathcal{F}_a(\bar{Y})]$ is analytic in $(\sup \operatorname{supp} \bar{\mu}_T, \infty)$, so both $\mathcal{S}$ and $\mathcal{S}'$ cannot have accumulating points in $(\sup \operatorname{supp} \bar{\mu}_T, \infty)$. Thus, in order to prove item (1), it suffices to prove that $\mathcal{S}, \mathcal{S}'$ are contained in a compact subset of $(\sup \operatorname{supp} \bar{\mu}_T, \infty)$. By the assumptions on $\mathcal{T}$ ((d) in (2.6)), we have

$$\lim_{a \searrow \sup \operatorname{supp} \bar{\mu}_T} -\mathbb{E}[\mathcal{F}_a(\bar{Y})] = -\infty,$$

hence $\mathcal{S}$ and $\mathcal{S}'$ are contained in $[x, \infty)$ for some $x > \sup \operatorname{supp} \bar{\mu}_T$. Also, we have the series expansion

$$-\mathbb{E}[\mathcal{F}_a(\bar{Y})] = -\frac{\mathbb{E}[\mathcal{T}(\bar{Y})]}{a} - \frac{\mathbb{E}[\mathcal{T}(\bar{Y})^2]}{a^2} + \mathcal{O}(a^{-3}), \quad \text{as } a \to \infty,$$

where $\mathbb{E}[\mathcal{T}(\bar{Y})^2] > 0$ by the assumption in (2.4). This already proves that $\mathcal{S}'$ is bounded, as $-\mathbb{E}[\mathcal{F}_a(\bar{Y})]$ converges to 0 as $a \to \infty$. Similarly, the same expansion implies that for large enough $x > \sup \operatorname{supp} \bar{\mu}_T$, we have

$$-\mathbb{E}[\mathcal{F}_a(\bar{Y})] \in \begin{cases} (0, \infty) & \text{if } \mathbb{E}[\mathcal{T}(\bar{Y})] < 0, \\ (-\infty, 0) & \text{if } \mathbb{E}[\mathcal{T}(\bar{Y})] \geq 0, \end{cases} \quad \forall a > x.$$

Thus, $\mathcal{S} \cap [x, \infty) = \emptyset$. This concludes item (1).

For item (2), we only need to notice that the right-hand side of (E.1) is a bijection between $\mathbb{R} \setminus (\inf \operatorname{supp} \bar{\mu}_\Sigma, \sup \operatorname{supp} \bar{\mu}_\Sigma)$ and $\mathbb{R} \setminus \{0\}$. Notice further that the right-hand side is analytic in ω with strictly positive derivative whenever ω is well defined:

$$\frac{\mathrm{d}}{\mathrm{d}\omega} \int_{\mathbb{R}} \frac{s}{s - \omega} \, \mathrm{d}\bar{\mu}_\Sigma = \int_{\mathbb{R}} \frac{s}{(s - \omega)^2} \, \mathrm{d}\bar{\mu}_\Sigma.$$

We now turn to item (3). Since the left-hand side of (E.1) is an analytic function of a, it immediately follows from analytic inverse function theorem that ω extends analytically to a neighborhood of $(\sup \operatorname{supp} \bar{\mu}_T, \infty) \setminus \mathcal{S}$. Similarly, for each $a > \sup \operatorname{supp} \mathcal{T}(\bar{Y})$ with $a \notin \mathcal{S} \cup \mathcal{S}'$, we find that $\tilde{\omega}(a) := 1/\omega(a)$ solves

$$\delta \int_{\mathbb{R}} \frac{t}{t - a} \, \mathrm{d}\bar{\mu}_T(t) = -\tilde{\omega}(a) \int_{\mathbb{R}} \frac{s}{1 - s\tilde{\omega}(a)} \, \mathrm{d}\bar{\mu}_\Sigma.$$

Defining $\tilde{\omega}(a) \equiv 0$ for $a \in \mathcal{S}$ and following the same reasoning as for ω, one easily finds that $\tilde{\omega}$ extends analytically to a neighborhood of $(\sup \operatorname{supp} \bar{\mu}_T, \infty) \setminus \mathcal{S}'$. By analytic continuation, ω extends to a meromorphic function on a neighborhood of $(\sup \operatorname{supp} \bar{\mu}_T, \infty)$ with poles at $\mathcal{S}$. From (E.1) we immediately find that the zeros of ω are exactly at $\mathcal{S}'$.

Finally, for item (4), note that by a trivial rescaling we have

$$-\omega(a) \int_{\mathbb{R}} \frac{t}{t-a} \, d\bar{\mu}_T(t) = \gamma(a),$$

which implies

$$\psi(a) = -a\omega(a) \int_{\mathbb{R}} \frac{t}{t-a} \, d\bar{\mu}_T(t), \quad a \notin S. \tag{E.3}$$

Using the definition of ω, we immediately have (E.2) from (E.3). Also, (E.3) already shows that ψ is a meromorphic function on a neighborhood of $(\sup \operatorname{supp} \bar{\mu}_T, \infty)$ by item (2), with possible poles at S. Hence, we only need to check that each $a \in S$ is a removable singularity for ψ. Recall that $\omega(z) \to \infty$ as $z \to a \in S$, so that by dominated convergence

$$\psi(z) = -\frac{z}{\delta} \int_{\mathbb{R}} \frac{s\omega(z)}{s - \omega(z)} \, d\bar{\mu}_\Sigma(s) = -\frac{z}{\delta} \int_{\mathbb{R}} \frac{s}{s/\omega(z) - 1} \, d\bar{\mu}_\Sigma(s) \to \frac{a}{\delta} \mathbb{E}[\bar{\Sigma}]. \quad \blacksquare$$

Lemma E.3. *We have*

$$\lim_{a \to \infty} \psi'(a) = \frac{\mathbb{E}[\bar{\Sigma}]}{\delta} = \lim_{\operatorname{Re} a \to \infty} \frac{\operatorname{Im} \psi(a)}{\operatorname{Im} a}, \tag{E.4}$$

where we identified ψ with its analytic extension. We also have

$$\lim_{a \to \infty} \psi(a) = \infty = \lim_{a \searrow \sup \operatorname{supp} \bar{\mu}_T} \psi(a). \tag{E.5}$$

In particular, the set of critical points of ψ is non-empty and bounded from above (as a subset of $\mathbb{R}$).

Proof. We compute the derivative of ψ as

$$\begin{aligned}
\delta\psi'(a) &= -\int_{\mathbb{R}} \frac{s\omega(a)}{s - \omega(a)} \, d\bar{\mu}_\Sigma(s) - a\omega'(a) \int_{\mathbb{R}} \frac{s^2}{(s - \omega(a))^2} \, d\bar{\mu}_\Sigma(s) \\
&= -\int_{\mathbb{R}} \frac{s\omega(a)}{s - \omega(a)} \, d\bar{\mu}_\Sigma(s) - a\delta \left(\int_{\mathbb{R}} \frac{s}{(s - \omega(a))^2} \, d\bar{\mu}_\Sigma(s) \right)^{-1} \\
&\quad \cdot \int_{\mathbb{R}} \frac{t}{(t-a)^2} \, d\bar{\mu}_T(t) \int_{\mathbb{R}} \frac{s^2}{(s - \omega(a))^2} \, d\bar{\mu}_\Sigma(s).
\end{aligned} \tag{E.6}$$

Furthermore, notice from item (2) of Lemma E.2 that $|\omega(a)| \to \infty$ as $a \to \infty$, so that the second term in (E.6) satisfies

$$\lim_{a \to \infty} \left(\int_{\mathbb{R}} \frac{s\omega(a)^2}{(s - \omega(a))^2} \, d\bar{\mu}_\Sigma(s) \right)^{-1} \int_{\mathbb{R}} \frac{s^2 \omega(a)^2}{(s - \omega(a))^2} \, d\bar{\mu}_\Sigma(s) \cdot \int_{\mathbb{R}} \frac{ta}{(t-a)^2} \, d\bar{\mu}_T(t)$$

$$= \frac{\mathbb{E}[\bar{\Sigma}^2]}{\mathbb{E}[\bar{\Sigma}]} \lim_{a \to \infty} \int_{\mathbb{R}} \frac{ta}{(t-a)^2} \, d\bar{\mu}_T(t) = 0.$$

Therefore, we conclude that the first equality in (E.4) holds as

$$\lim_{a\to\infty} \psi'(a) = \frac{1}{\delta} \lim_{a\to\infty} \int_{\mathbb{R}} \frac{-s\omega(a)}{s-\omega(a)} \,\mathrm{d}\bar{\mu}_\Sigma(s) = \frac{1}{\delta}\mathbb{E}[\bar{\Sigma}].$$

The second equality can be proved analogously, except that the following identity replaces (E.6):

$$\delta \frac{\operatorname{Im}\psi(a)}{\operatorname{Im}a} = -\operatorname{Re}\left[\int_{\mathbb{R}} \frac{s\omega(a)}{s-\omega(a)} \,\mathrm{d}\bar{\mu}_\Sigma(s)\right] - \delta\operatorname{Re}[a]\left(\int_{\mathbb{R}} \frac{s}{|s-\omega(a)|^2} \,\mathrm{d}\bar{\mu}_\Sigma(s)\right)^{-1}$$
$$\cdot \int_{\mathbb{R}} \frac{t}{|t-a|^2} \,\mathrm{d}\bar{\mu}_T(t) \int_{\mathbb{R}} \frac{s^2}{|s-\omega(a)|^2} \,\mathrm{d}\bar{\mu}_\Sigma(s),$$

where we used

$$\frac{\operatorname{Im}\omega(a)}{\operatorname{Im}a} = \delta\left(\int_{\mathbb{R}} \frac{s}{|s-\omega(a)|^2} \,\mathrm{d}\bar{\mu}_\Sigma(s)\right)^{-1} \int_{\mathbb{R}} \frac{t}{|t-a|^2} \,\mathrm{d}\bar{\mu}_T(t),$$

from (E.1).

Notice that the first equality in (E.5) follows from the first equality in (E.4). For the second equality in (E.5), recall from assumption (d) in (2.6) that

$$\lim_{a\searrow\text{sup supp}\,\bar{\mu}_T} \int_{\mathbb{R}} \frac{t}{t-a} \,\mathrm{d}\bar{\mu}_T(t) = -\infty,$$

which implies

$$\lim_{a\searrow\text{sup supp}\,\bar{\mu}_T} \omega(a) = \text{sup supp}\,\bar{\mu}_\Sigma$$

via item (2) of Lemma E.2. Plugging these in the definition of ψ in (E.2) and using sup supp $\bar{\mu}_T > 0$ prove $\psi(a) \to \infty$. ∎

E.2. Complex analytic characterization of $\bar{\mu}_{\hat{D}}$

Lemma E.4 ([98, Theorem 1.2.1]). *Let $m_{\bar{\mu}_{\hat{D}}}$ denote the Stieltjes transform of the limiting eigenvalue distribution $\bar{\mu}_{\hat{D}}$. For each $z \in \mathbb{H} := \{z \in \mathbb{C} : \Im(z) > 0\}$, $m = m_{\bar{\mu}_{\hat{D}}}(z)$ is characterized as the unique solution (m, m_1, m_2) of the following system of equations:*

$$\begin{cases} -zm = (1-\delta) + \delta \int_{\mathbb{R}} 1/(1+m_1 t) \,\mathrm{d}\bar{\mu}_T(t), \\ -zm = \int_{\mathbb{R}} 1/(1+m_2 s) \,\mathrm{d}\bar{\mu}_\Sigma(s), \\ -zm = 1 + \delta z m_1 m_2, \end{cases} \tag{E.7}$$

subject to the constraint $m, m_1, zm_2 \in \mathbb{H}$. All of m, m_1, m_2 are analytic in $\mathbb{H}$ as a function of z.

We adopt the notation $m(\bar{z}) = \bar{m}(z)$ and $m_i(\bar{z}) = \overline{m_i(z)}$ ($i \in \{1, 2\}$). The major difference from the case of positive $\mathcal{T}$ is that m_2 might not be in $\mathbb{H}$; still the second equation in (E.7) is well defined as $m_2(z) \in \{z^{-1}w : w \in \mathbb{H}\} \subset \mathbb{C} \setminus (-\infty, 0]$ (cf. when $\mathcal{T}$ is positive then $m_i \in \mathbb{H}$ and $z m_i \in \mathbb{H}$ for both $i \in \{1, 2\}$). Alternatively, using the last equation in (E.7) to substitute m in the first two equations, we may write the system of two equations for m_1, m_2:

$$\begin{cases} -z m_1 = \frac{1}{\delta} \int_{\mathbb{R}} s/(1 + m_2 s) \, \mathrm{d}\bar{\mu}_\Sigma(s), \\ -z m_2 = \int_{\mathbb{R}} t/(1 + m_1 t) \, \mathrm{d}\bar{\mu}_T(t). \end{cases} \tag{E.8}$$

For later purposes, we define for all $z, w \in \mathbb{C} \setminus \mathbb{R}$,

$$I_1(z, w) := \int_{\mathbb{R}} \frac{t^2}{(1 + m_1(z)t)(1 + m_1(w)t)} \, \mathrm{d}\bar{\mu}_T(t),$$

$$I_2(z, w) := \int_{\mathbb{R}} \frac{s^2}{(1 + m_2(z)s)(1 + m_2(w)s)} \, \mathrm{d}\bar{\mu}_\Sigma(s),$$

so that $I_1(z, \bar{z})$ and $I_2(z, \bar{z})$ are positive since $m_i(\bar{z}) = \overline{m_i(z)}$. Note also that

$$|z m_1(z)| \leq \delta^{-1} I_2(z, \bar{z})^{1/2}, \quad |z m_2(z)| \leq I_1(z, \bar{z})^{1/2}, \tag{E.9}$$

by Cauchy–Schwarz.

Lemma E.5. *For all $z \in \mathbb{H}$,*

$$\frac{1}{\delta |z|^2} I_1(z, \bar{z}) I_2(z, \bar{z}) < 1. \tag{E.10}$$

Consequently,

$$|m_1(z)|^2 I_1(z, \bar{z}) < \frac{1}{\delta}, \quad |m_2(z)|^2 I_2(z, \bar{z}) < \delta. \tag{E.11}$$

Proof. Dividing the first line of (E.8) by z and then taking imaginary parts, we get

$$m_1(z) = \frac{1}{\delta} \int_{\mathbb{R}} \frac{s}{-z(1 + m_2(z)s)} \, \mathrm{d}\bar{\mu}_\Sigma(s) = \frac{1}{\delta} \int_{\mathbb{R}} \frac{sz + s^2 z m_2(z)}{|z|^2 |1 + m_2(z)s|^2} \, \mathrm{d}\bar{\mu}_\Sigma(t). \tag{E.12}$$

Similarly taking the imaginary part of the second line of (E.8) gives

$$z m_2(z) = -\int_{\mathbb{R}} \frac{t}{1 + m_1(z)t} \, \mathrm{d}\bar{\mu}_T(t) = \int_{\mathbb{R}} \frac{t^2 m_1(z)}{|1 + m_1(z)t|^2} \, \mathrm{d}\bar{\mu}_T(t). \tag{E.13}$$

Combining (E.12) and (E.13), we obtain

$$\delta m_1(z) = \int_{\mathbb{R}} \frac{sz}{|z|^2 |1 + m_2(z)s|^2} \, \mathrm{d}\bar{\mu}_\Sigma(s)$$
$$+ \frac{m_1(z)}{|z|^2} \left(\int_{\mathbb{R}} \frac{t^2}{|1 + m_1(z)t|^2} \, \mathrm{d}\bar{\mu}_T(t) \right) \left(\int_{\mathbb{R}} \frac{s^2}{|1 + m_2(z)s|^2} \, \mathrm{d}\bar{\mu}_\Sigma(s) \right). \tag{E.14}$$

Since $\operatorname{Im} m_1(z)$ and the first term on the right-hand side of (E.14) are positive for all $z \in \mathbb{H}$, we have proved (E.10):

$$\frac{1}{\delta |z|^2} \left(\int_{\mathbb{R}} \frac{t^2}{|1 + m_1(z)t|^2} \, \mathrm{d}\bar{\mu}_T(t) \right) \left(\int_{\mathbb{R}} \frac{s^2}{|1 + m_2(z)s|^2} \, \mathrm{d}\bar{\mu}_{\Sigma}(s) \right) < 1, \quad \forall z \in \mathbb{H}.$$

For (E.11), we only need to notice from (E.10) and (E.9) that

$$|m_1|^2 I_1(z, \bar{z}) \le \frac{1}{\delta^2 |z|^2} I_1(z, \bar{z}) I_2(z, \bar{z}) < \frac{1}{\delta},$$

and the second line in (E.11) follows similarly. ∎

Note also that (E.10) implies for all $z \in \mathbb{H}$ that

$$|zm(z) + 1| \le \frac{\delta}{|z|} |zm_1(z)| |zm_2(z)| \le \frac{1}{|z|} \sqrt{I_1(z, \bar{z}) I_2(z, \bar{z})} \le \sqrt{\delta}, \qquad \text{(E.15)}$$

where we used the third line of (E.7) in the first, (E.9) in the second, and (E.10) in the last inequality.

Lemma E.6. *Let $\mathcal{D} \subset \mathbb{H}$ be bounded. Then, there exists a constant $K > 0$ depending only on $\mathcal{D}$, $\bar{\mu}_{\Sigma}$, and $\bar{\mu}_T$ such that*

$$|zm_1(z)| \le K, \quad |zm_2(z)| \le K, \qquad \forall z \in \mathcal{D}.$$

Proof. We only consider $|zm_1(z)|$, and the same argument applies to $|zm_2(z)|$. The proof is by contradiction. Suppose that there exists a sequence z_k in $\mathcal{D}$ such that

$$|z_k m_1(z_k)| \to \infty.$$

Then by combining (E.15) with the third equation in (E.7), we have $|m_2(z_k)| \to 0$. Therefore, by dominated convergence (together with $\sup \operatorname{supp} \bar{\mu}_{\Sigma} < \infty$), we have

$$-\delta \lim_{k \to \infty} z_k m_1(z_k) = \lim_{k \to \infty} \int_{\mathbb{R}} \frac{s}{1 + m_2(z_k)s} \, \mathrm{d}\bar{\mu}_{\Sigma}(s) = \int_{\mathbb{R}} s \, \mathrm{d}\bar{\mu}_{\Sigma}(s) \in \mathbb{R},$$

which gives a contradiction to $|z_k m_1(z_k)| \to \infty$. ∎

Lemma E.7. *For all $z \in \mathbb{H}$, we have*

$$0 < (\inf \operatorname{supp} \bar{\mu}_{\Sigma}) \le \delta \frac{\operatorname{Im} m_1(z)}{\operatorname{Im} m(z)} \le (\sup \operatorname{supp} \bar{\mu}_{\Sigma}). \qquad \text{(E.16)}$$

For each bounded $\mathcal{D} \subset \mathbb{H}$, there exists a constant K_1 depending only on $\mathcal{D}, \bar{\mu}_{\Sigma},$ and $\bar{\mu}_T$ such that

$$\operatorname{Im}(zm_2(z)) \le K_1 \operatorname{Im} m_1(z), \quad z \in \mathcal{D}. \qquad \text{(E.17)}$$

Proof. To see (E.16), note that the second line of (E.7) implies

$$\operatorname{Im} m(z) = \int_{\mathbb{R}} \operatorname{Im}\left[\frac{1}{-z(1 + m_2(z)s)}\right] d\bar{\mu}_{\Sigma}(s). \qquad (\text{E.18})$$

Comparing (E.18) with (E.12) proves (E.16).

For (E.17), we recall from (E.13) and (E.10) that

$$\operatorname{Im} z m_2(z) = \operatorname{Im} m_1(z) \cdot I_1(z, \bar{z}) \leq \operatorname{Im} m_1(z) \cdot \frac{\delta |z|^2}{I_2(z, \bar{z})}.$$

By definition of $I_2(z, \bar{z})$, we have

$$\frac{|z|^2}{I_2(z, \bar{z})} = \left(\int_{\mathbb{R}} \frac{s^2}{|z + z m_2(z)s|^2} d\bar{\mu}_{\Sigma}(s)\right)^{-1}$$

$$\leq 2\left(\left[(\sup \operatorname{supp} \bar{\mu}_{\Sigma}) \cdot |z m_2(z)|\right]^2 + |z|^2\right)\left(\int_{\mathbb{R}} s^2 d\bar{\mu}_{\Sigma}(s)\right)^{-1}. \qquad (\text{E.19})$$

Since $\mathcal{D}$ is bounded, the right-hand side of (E.19) is bounded by a constant for all $z \in \mathcal{D}$. This proves (E.17). ∎

Proposition E.8. *The following assertions hold:*

(1) *There exist two finite measures ν_1, ν_2 on $\mathbb{R}$ such that the following holds: for all $z \in \mathbb{H}$, we have*

$$\int_{\mathbb{R}} \frac{1}{x - z} d\nu_1(x) = m_1(z), \qquad\qquad \nu_1(\mathbb{R}) = \frac{\mathbb{E}[\bar{\Sigma}]}{\delta},$$

$$\int_{\mathbb{R}} \frac{1}{x - z} d\nu_2(x) = z m_2(z) + \int_{\mathbb{R}} t \, d\bar{\mu}_T(t), \quad \nu_2(\mathbb{R}) = \frac{\mathbb{E}[\bar{\Sigma}]}{\delta} \mathbb{E}[\mathcal{T}(\bar{Y})^2]. \qquad (\text{E.20})$$

Consequently, we have

$$\operatorname{supp} \nu_1 = \operatorname{supp} \bar{\mu}_{\hat{D}}, \quad \operatorname{supp} \nu_2 \subset \operatorname{supp} \bar{\mu}_{\hat{D}}, \qquad (\text{E.21})$$

so that m_1 and m_2 are respectively analytic and meromorphic functions on $\mathbb{R} \setminus \operatorname{supp} \bar{\mu}_{\hat{D}}$.

(2) *For all $x > \lambda^\circ$, we have*

$$-\frac{1}{m_1(x)} \in (\sup \operatorname{supp} \bar{\mu}_T, \infty),$$

$$-\frac{1}{m_2(x)} \in \left(\mathbb{R} \cup \{\infty\}\right) \setminus (\inf \operatorname{supp} \bar{\mu}_{\Sigma}, \sup \operatorname{supp} \bar{\mu}_{\Sigma}), \qquad (\text{E.22})$$

$$\limsup_{z \to x, \, z \in \mathbb{H}} \frac{1}{\delta |z|^2} I_1(z, \bar{z}) I_2(z, \bar{z}) < 1,$$

where we used the convention $1/0 = \infty$ in the second assertion.

Proof. We start with the proof of item (1). First, notice that once (E.20) is proved, (E.21) immediately follows from Lemma E.7 and Stieltjes inversion. In order to prove the first identity in (E.20), since m_1 is an analytic self-map of $\mathbb{H}$, by Nevanlinna–Pick representation theorem it suffices to check

$$\limsup_{\eta \to \infty} \eta |m_1(i\eta)| < \infty. \tag{E.23}$$

Suppose the contrary, so that there exists a sequence $\eta_k \to \infty$ with $\eta_k |m_1(i\eta_k)| \to \infty$. Then by (E.15), we find that $|m_2(i\eta_k)| \to 0$. On the other hand by (E.8), we have

$$-i\eta m_1(i\eta) = \frac{1}{\delta} \int_{\mathbb{R}} \frac{s}{1 + m_2(i\eta)s} \, d\bar{\mu}_\Sigma(s), \tag{E.24}$$

so that the dominated convergence theorem (with $\|\Sigma\|_2 = \mathcal{O}(1)$) leads to a contradiction as

$$\lim_{k \to \infty} \eta_k |m_1(i\eta_k)| = \frac{1}{\delta} \lim_{k \to \infty} \left| \int_{\mathbb{R}} \frac{s}{1 + m_2(i\eta_k)s} \, d\bar{\mu}_\Sigma(s) \right| = \frac{1}{\delta} \int_{\mathbb{R}} s \, d\bar{\mu}_\Sigma(s).$$

Thus we have proved the first line of (E.20).

Next, we prove the corresponding representation for $zm_2(z)$, the second line of (E.20). As before, it suffices to prove

$$\limsup_{\eta \to \infty} \eta \left| i\eta m_2(i\eta) + \int_{\mathbb{R}} t \, d\bar{\mu}_T(t) \right| < \infty.$$

To this end, we use (E.8) to write

$$z \left(zm_2(z) + \int_{\mathbb{R}} t \, d\bar{\mu}_T(t) \right) = zm_1(z) \int_{\mathbb{R}} \frac{t^2}{1 + m_1(z)t} \, d\bar{\mu}_T(t). \tag{E.25}$$

Taking the limit along $z = i\eta \to i\infty$, by (E.20) we have

$$m_1(z) \to 0 \quad \text{and} \quad zm_1(z) \to -\nu_1(\mathbb{R})$$

(note that $\nu_1(\mathbb{R})$ is finite due to (E.23)), so that

$$\lim_{\eta \to \infty} i\eta \left(i\eta m_2(i\eta) + \int_{\mathbb{R}} t \, d\bar{\mu}_T(t) \right) = -\nu_1(\mathbb{R}) \int_{\mathbb{R}} t^2 \, d\bar{\mu}_T(t).$$

Finally, given the two representations in (E.20), we have

$$m_1(i\eta), m_2(i\eta) \to 0 \text{ as } \eta \to \infty.$$

Then $\nu_1(\mathbb{R})$ and $\nu_2(\mathbb{R})$ can be computed by taking the limits of (E.24) and (E.25) as $z = i\eta \to i\infty$. This completes the proof of item (1).

Now we prove item (2). Notice that m_1 is analytic, negative-valued, and increasing on $(\lambda°, \infty)$, and that $\lim_{x \to \infty} m_1(x) = 0$. Therefore, the image of the half line $(\lambda°, \infty)$ under $x \mapsto -1/m_1(x)$ is again a half-line (y_0, ∞) for some $y_0 > 0$. Next, notice from (E.11) that for all $x \in \mathbb{R}$,

$$\limsup_{z \to x, z \in \mathbb{H}} |m_1(z)|^2 I_1(z, \overline{z}) = \limsup_{z \to x, z \in \mathbb{H}} \int_{\mathbb{R}} \frac{t^2}{|t - (-1/m_1(z))|^2} \, d\overline{\mu}_T(t) < \frac{1}{\delta}. \quad \text{(E.26)}$$

On the other hand, by the assumptions on $\mathcal{T}$ (see (d) in (2.6)) and Cauchy–Schwarz, there exists an $\varepsilon > 0$ so that

$$\lim_{w \to y, w \in \mathbb{H}} \int_{\mathbb{R}} \frac{t^2}{|t - w|^2} \, d\overline{\mu}_T(t) = \int_{\mathbb{R}} \frac{t^2}{|t - y|^2} \, d\overline{\mu}_T(t) > \frac{1}{\delta} \quad \text{(E.27)}$$

for all $y \in (\sup \operatorname{supp} \overline{\mu}_T, \sup \operatorname{supp} \overline{\mu}_T + \varepsilon)$. Combining (E.26) and (E.27), we conclude that (y_0, ∞) does not intersect with $(\sup \operatorname{supp} \overline{\mu}_T, \sup \operatorname{supp} \overline{\mu}_T + \varepsilon)$, so that

$$y_0 \geq \sup \operatorname{supp} \overline{\mu}_T + \varepsilon.$$

This proves the first assertion of item (2).

The proof of the second assertion in item (2) follows similar lines, except that we view $x \mapsto -1/m_2(x)$ as an analytic (instead of meromorphic) function mapping into the Riemann sphere $\mathbb{C} \cup \{\infty\}$. Consequently, the closure of the image of $(\lambda°, \infty)$ under $z \mapsto -1/m_2(z)$ is a connected real interval in the Riemann sphere; or equivalently, it is the image of a closed connected arc in the unit circle under stereographic projection. Next, notice from the assumptions on $\overline{\Sigma}$ (see (b) in (2.5)) that there exists an $\varepsilon > 0$ so that

$$\lim_{w \to y, w \in \mathbb{H}} \int_{\mathbb{R}} \frac{s^2}{|s - w|^2} \, d\overline{\mu}_{\Sigma}(s) > \delta$$

for all $y \in (\inf \operatorname{supp} \overline{\mu}_{\Sigma} - \varepsilon, \inf \operatorname{supp} \overline{\mu}_{\Sigma}) \cup (\sup \operatorname{supp} \overline{\mu}_{\Sigma}, \sup \operatorname{supp} \overline{\mu}_{\Sigma} + \varepsilon)$. Therefore, (E.11) implies that the image of $(\lambda°, \infty)$ under $x \mapsto -1/m_2(x)$ does not intersect with the two segments of length ε, while containing ∞ in its closure since $m_2(x) \to 0$ as $x \to \infty$. This proves the second assertion of item (2).

For the final assertion of item (2), recall from (E.14) that for all $z \in \mathbb{H}$,

$$1 - \frac{1}{\delta |z|^2} I_1(z, \overline{z}) I_2(z, \overline{z})$$

$$= \frac{\operatorname{Im} z}{\delta \operatorname{Im} m_1(z)} \int_{\mathbb{R}} \frac{s}{|z|^2 |1 + m_2(z) s|^2} \, d\overline{\mu}_{\Sigma}(s)$$

$$= \left(\delta \int_{\mathbb{R}} \frac{1}{|y - z|^2} \, d\nu_1(y) \right)^{-1} \int_{\mathbb{R}} \frac{s}{|z|^2 |1 + m_2(z) s|^2} \, d\overline{\mu}_{\Sigma}(s),$$

where we used (E.20) in the second equality. Taking the limit $z \to x > \lambda^\circ$, we have

$$1 - \limsup_{z \to x, \, z \in \mathbb{H}} \frac{1}{\delta |z|^2} I_1(z, \overline{z}) I_2(z, \overline{z})$$

$$= \left(\delta \int_{\mathbb{R}} \frac{1}{|y - x|^2} \, \mathrm{d}\nu_1(y) \right)^{-1} \int_{\mathbb{R}} s \left(x^2 + s^2 \limsup_{z \to x, \, z \in \mathbb{H}} |z m_2(z)|^2 \right)^{-1} \mathrm{d}\overline{\mu}_\Sigma(s) > 0,$$

where we used Fatou's lemma in the first equality and Lemma E.6 in the last inequality. This concludes the proof of Proposition E.8. ∎

E.3. Proofs of Lemmas 5.5 and E.1

Proof of Lemma 5.5 given Lemma E.1. Notice that since a° is the largest critical point of ψ and $\lim_{a \to \infty} \psi'(a) > 0$, we find that $\psi'(a) > 0$ for all $a \in (a^\circ, \infty)$, i.e. ψ is strictly increasing on $[a^\circ, \infty)$.

Next, we prove $\psi(a^\circ) \leq \lambda^\circ$. Note from the contrapositive of item (1) of Lemma E.1 that if $a > \sup \mathrm{supp}\, \overline{\mu}_T$ and $\psi'(a) \leq 0$, then there exists an $\widetilde{a} \geq a$ such that $\psi(\widetilde{a}) \leq \lambda^\circ$. We may apply this to the largest critical point a° since $\psi'(a^\circ) = 0$, so that $\psi(\widetilde{a}) \leq \lambda^\circ$ for some $\widetilde{a} \geq a^\circ$. As ψ is increasing in $[a^\circ, \infty)$, we conclude $\psi(a^\circ) \leq \psi(\widetilde{a}) \leq \lambda^\circ$

Conversely, item (2) of Lemma E.1 implies $(\psi(a^\circ), \infty) \cap \mathrm{supp}\, \overline{\mu}_{\widehat{D}} = \emptyset$, so that $\lambda^\circ \leq \psi(a^\circ)$. Therefore, we have $\psi(a^\circ) = \lambda^\circ$. ∎

Proof of item (1) of Lemma E.1. Let $a \in (\sup \mathrm{supp}\, \overline{\mu}_T, \infty)$ satisfy the assumption of item (1) of Lemma E.1, that is, $\psi(\widetilde{a}) > \lambda^\circ$ for all $\widetilde{a} \geq a$. First of all, we prove that there exists a complex neighborhood U of $[a, \infty)$ such that

$$w = -1/m_1(\psi(w)), \quad \omega(w) = -1/m_2(\psi(w)), \qquad \forall w \in U. \tag{E.28}$$

Here we remark that $\psi(a) > \lambda^\circ$ by assumption, so that all four functions of w in (E.28) are well defined by Proposition E.8; those in the first and second equalities are analytic and meromorphic, respectively.

Recall from Lemma E.3 that for large enough $\widetilde{a} > a$, there exists a neighborhood V of $\widetilde{a}$ so that $\mathrm{Im}\, \psi(w)/\mathrm{Im}\, w > 0$ for every $w \in V$. Then, it also follows that, for each $w \in V \cap \mathbb{H}$,

$$\mathrm{Im}\left[-\frac{\psi(w)}{\omega(w)} \right] = \mathrm{Im}\left[\int_{\mathbb{R}} \frac{t w}{t - w} \, \mathrm{d}\overline{\mu}_T(t) \right] = \mathrm{Im}\, w \int_{\mathbb{R}} \frac{t^2}{|t - w|^2} \, \mathrm{d}\overline{\mu}_T(t) > 0. \tag{E.29}$$

Also notice that the triple $(\psi(w), -1/w, -1/\omega(w))$ satisfies the same system of equations as in (E.8):

$$-\frac{\psi(w)}{w} = \frac{1}{\delta} \int_{\mathbb{R}} \frac{s \omega(w)}{s - \omega(w)} \, \mathrm{d}\overline{\mu}_\Sigma(s) = -\frac{1}{\delta} \int_{\mathbb{R}} \frac{s}{1 + s \cdot (-\omega(w))^{-1}} \, \mathrm{d}\overline{\mu}_\Sigma(s),$$

$$\frac{\psi(w)}{\omega(w)} = -\int_{\mathbb{R}} \frac{t w}{t - w} \, \mathrm{d}\overline{\mu}_T(t) = -\int_{\mathbb{R}} \frac{t}{1 + t \cdot (-w^{-1})} \, \mathrm{d}\overline{\mu}_T(t). \tag{E.30}$$

Therefore, by the uniqueness of the solution of (E.8), we conclude

$$\big(\psi(w), -1/w, -1/\omega(w)\big) = \big(\psi(w), m_1(\psi(w)), m_2(\psi(w))\big), \quad w \in V \cap \mathbb{H}. \quad \text{(E.31)}$$

By Proposition E.8 and the assumption of item (1), in both sides of (E.31) are mero-morphic functions defined on a neighborhood of $[a, \infty)$, so that the identity holds in the whole (connected) neighborhood.

We now prove $\psi'(a) > 0$, provided $a \notin S \cup S'$. Recall from (E.6) that

$$\delta\psi'(a) = \left(\int_{\mathbb{R}} \frac{s}{(s-\omega(a))^2}\, \mathrm{d}\bar\mu_\Sigma(s)\right)^{-1}$$
$$\cdot \left[-\delta \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{t}{t-a} \frac{s\omega(a)}{(s-\omega(a))^2} + \frac{ta}{(t-a)^2} \frac{s^2}{(s-\omega(a))^2}\, \mathrm{d}\bar\mu_\Sigma(s)\, \mathrm{d}\bar\mu_T(t)\right]. \quad \text{(E.32)}$$

Note that the second line in (E.32) can be written as

$$-\delta \int_{\mathbb{R}} \int_{\mathbb{R}} \left[-\frac{t}{t-a} \frac{s}{s-\omega(a)} + \frac{t^2}{(t-a)^2} \frac{s^2}{(s-\omega(a))^2}\right] \mathrm{d}\bar\mu_\Sigma(s)\, \mathrm{d}\bar\mu_T(t)$$
$$= \frac{\delta^2 \psi(a)^2}{a^2 \omega(a)^2} - \delta \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{t^2}{(t-a)^2} \frac{s^2}{(s-\omega(a))^2}\, \mathrm{d}\bar\mu_\Sigma(s)\, \mathrm{d}\bar\mu_T(t).$$

Then, we use (E.28) for $w = a$ to substitute a and $\omega(a)$ in (E.32) to obtain

$$\psi'(a) = \frac{\delta\psi(a)^2}{a^2\omega(a)^2} \left(\int_{\mathbb{R}} \frac{s}{(s-\omega(a))^2}\, \mathrm{d}\bar\mu_\Sigma(s)\right)^{-1}$$
$$\cdot \left(1 - \frac{1}{\delta\psi(a)^2} I_1\big(\psi(a), \psi(a)\big) I_2\big(\psi(a), \psi(a)\big)\right) > 0, \quad \text{(E.33)}$$

where we used $0 < |m_2(\psi(a))|, |\psi(a)| < \infty$ for $a \neq S \cup S'$ and (E.22).

It only remains to prove $\psi'(a) > 0$ for $a \in S \cup S'$. Since S and S' are both finite, we may consider a sequence $\tilde{a}_k > a$ such that $\tilde{a}_k \notin S \cup S'$ and $\tilde{a}_k \to a$. Since ψ is analytic at a and the second line of (E.33) is strictly positive by Proposition E.8, is suffices to prove

$$\lim_{k\to\infty} \frac{\psi(\tilde{a}_k)^2}{\omega(\tilde{a}_k)^2} \left(\int_{\mathbb{R}} \frac{s}{(s-\omega(\tilde{a}_k))^2}\, \mathrm{d}\bar\mu_\Sigma(s)\right)^{-1} > 0.$$

If $a \in S$ so that $\omega(\tilde{a}_k) \to \infty$, we have

$$\lim_{k\to\infty} \frac{\psi(\tilde{a}_k)^2}{\omega(\tilde{a}_k)^2} \left(\int_{\mathbb{R}} \frac{s}{(s-\omega(\tilde{a}_k))^2}\, \mathrm{d}\bar\mu_\Sigma(s)\right)^{-1}$$
$$= \psi(a)^2 \lim_{k\to\infty} \left(\int_{\mathbb{R}} s \left(\frac{\omega(\tilde{a}_k)}{s-\omega(\tilde{a}_k)}\right)^2 \mathrm{d}\bar\mu_\Sigma(s)\right)^{-1}$$
$$= \frac{\psi(a)^2}{\mathbb{E}[\bar\Sigma]} > 0,$$

where in the last inequality we used $a \in S$ implies $a \notin S'$, which in turn gives $\psi(a) \neq 0$. Finally, for $a \in S'$, we use $\omega(\tilde{a}_k) \to 0$ to write

$$
\begin{aligned}
\lim_{k \to \infty} \frac{\psi(\tilde{a}_k)^2}{\omega(\tilde{a}_k)^2} \left(\int_{\mathbb{R}} \frac{s}{(s - \omega(\tilde{a}_k))^2} \, d\bar{\mu}_\Sigma(s) \right)^{-1} &= \frac{1}{\mathbb{E}[\bar{\Sigma}^{-1}]} \lim_{k \to \infty} \frac{\psi(\tilde{a}_k)^2}{\omega(\tilde{a}_k)^2} \\
&= \frac{a^2}{\delta^2 \mathbb{E}[\bar{\Sigma}^{-1}]} \lim_{k \to \infty} \left(\int_{\mathbb{R}} \frac{s}{s - \omega(\tilde{a}_k)} \, d\bar{\mu}_\Sigma(s) \right)^2 = \frac{a^2}{\delta^2 \mathbb{E}[\bar{\Sigma}^{-1}]} > 0,
\end{aligned}
$$

where we used the definition of ψ in the second equality and inf supp $\bar{\mu}_\Sigma > 0$ in the last inequality. This concludes the proof of item (1) of Lemma E.1. ∎

Proof of item (2) *of Lemma* E.1. Since $\psi'(a) > 0$, there exist small neighborhoods U and V respectively of a and $\psi(a)$ and an analytic inverse function $\psi^{-1} : V \to U$ of ψ. We first prove that

$$
\left(z, \frac{-1}{\psi^{-1}(z)}, \frac{-1}{\omega(\psi^{-1}(z))} \right) = (z, m_1(z), m_2(z)), \quad \forall z \in V \cap \mathbb{H}. \tag{E.34}
$$

Following (E.30), we easily find that $(z, -1/\psi^{-1}(z), -1/\omega(\psi^{-1}(z)))$ satisfies (E.8). Also, there is an open subset $V' \subset V \cap \mathbb{H}$ so that Im $\psi^{-1}(z) > 0$ for all $z \in V'$; to see this, we write

$$
\begin{aligned}
\operatorname{Im} \psi^{-1}(z) &= \operatorname{Im}\left[(\psi^{-1})'(\psi(a)) \cdot (z - \psi(a)) \right] + \mathcal{O}\left(|z - \psi(a)|^2 \right) \\
&= \frac{1}{\psi'(a)} \operatorname{Im} z + \mathcal{O}\left(|z - \psi(a)|^2 \right).
\end{aligned}
$$

Hence, it suffices to take $V' = \{z : |z - \psi(a)| < 2 \operatorname{Im} z < r\}$ with small enough $r > 0$ in order to have $\psi^{-1}(V') \subset \mathbb{H}$. Then, by (E.29) it also follows that

$$
\operatorname{Im}\left[-\frac{z}{\omega(\psi^{-1}(z))} \right] > 0.
$$

As in the proof of item (1) of Lemma E.1, the uniqueness of the solution of (E.8) implies (E.34) for $z \in V'$. Finally, the conclusion extends to $V \cap \mathbb{H}$ by analytic continuation.

Since ψ maps (sup supp $\bar{\mu}_T, \infty$) to $\mathbb{R}$, its inverse function ψ^{-1} is real-valued on $V \cap \mathbb{R}$. Hence, it follows

$$
\lim_{\eta \to 0} \operatorname{Im} m_1(x + i\eta) = \lim_{\eta \to 0} \operatorname{Im}\left[-\frac{1}{\psi^{-1}(x + i\eta)} \right] = 0, \quad x \in V \cap \mathbb{R}.
$$

Then, applying Stieltjes inversion to (E.20), we have supp $\nu_1 \cap V = \emptyset$. And finally, by (E.20), we conclude supp $\bar{\mu}_{\hat{D}} \cap V = \emptyset$, so that $\psi(a) \notin$ supp $\bar{\mu}_{\hat{D}}$. This completes the proof of item (2) in Lemma E.1. ∎

F. Performance of the whitened spectral estimator

In this section, we characterize the limiting overlap of the whitened spectral estimator, whose definition we recall from (4.1):

$$\beta_\triangle^{\text{spec}}(y, X, \Sigma) := \Sigma^{-1/2} v_1(D_\triangle), \qquad (\text{F.1})$$

where

$$D_\triangle := \sum_{i=1}^{n} (\Sigma^{-1/2} x_i)(\Sigma^{-1/2} x_i)^\top \mathcal{T}(y_i)$$

$$= \Sigma^{-1/2} X^\top T X \Sigma^{-1/2} = \widetilde{X}^\top T \widetilde{X} = \Sigma^{-1/2} D \Sigma^{-1/2}.$$

As discussed in Section 4, one can think of $\Sigma^{1/2}\beta^*$ as an auxiliary parameter in the model $y = q(\widetilde{X}\Sigma^{1/2}\beta^*, \varepsilon)$ with design matrix $\widetilde{X}$. Therefore, the top eigenvector of

$$D_\triangle = \widetilde{X}^\top \text{diag}\big(\mathcal{T}\big(q(\widetilde{X}\Sigma^{1/2}\beta^*, \varepsilon)\big)\big)\widetilde{X}$$

estimates $\Sigma^{1/2}\beta^*$ and $\Sigma^{-1/2}v_1(D_\triangle)$ estimates β^*. We highlight that computing this spectral estimator requires knowledge of Σ.

As before, our results concerning $\beta_\triangle^{\text{spec}}$ are expressed in terms of a few functions and parameters. Define

$$\varphi_\triangle, \psi_\triangle, \zeta_\triangle \colon \big(\sup \text{supp}(\mathcal{T}(\overline{Y})), \infty\big) \to \mathbb{R}, a_\triangle^\circ \in \big(\sup \text{supp}(\mathcal{T}(\overline{Y})), \infty\big)$$

as

$$\varphi_\triangle(a) = \frac{a\delta}{\mathbb{E}[\overline{\Sigma}]}\mathbb{E}\big[\overline{G}^2 \mathcal{F}_a(\overline{Y})\big], \qquad \psi_\triangle(a) = a\Big(\frac{1}{\delta} + \mathbb{E}[\mathcal{F}_a(\overline{Y})]\Big),$$

$$a_\triangle^\circ = \underset{a \in (\sup \text{supp}(\mathcal{T}(\overline{Y})), \infty)}{\arg\min} \psi_\triangle(a), \quad \zeta_\triangle(a) = \psi_\triangle\big(\max\{a, a_\triangle^\circ\}\big),$$

where $\mathcal{F}_a$ is given in (2.3), and $a_\triangle^* \in (\sup \text{supp}(\mathcal{T}(\overline{Y})), \infty)$ as the unique solution to

$$\zeta_\triangle(a_\triangle^*) = \varphi_\triangle(a_\triangle^*).$$

Both $a_\triangle^\circ$ and $a_\triangle^*$ are uniquely defined, as shown in [55, Theorem 2.1 (1)] and [62, Lemma 2 (1)]. In fact,

$$\sqrt{\frac{\delta}{\mathbb{E}[\overline{\Sigma}]}}\, \overline{G} \sim \mathcal{N}(0, 1),$$

so our functions $\varphi_\triangle, \psi_\triangle, \zeta_\triangle$ match φ, ψ, ζ in [55] by taking κ in [55] to be $\sqrt{\mathbb{E}[\overline{\Sigma}]/\delta}$. The formula of the asymptotic overlap $\eta_\triangle$ is:

$$\eta_\triangle := \left(\frac{1 - \delta\mathbb{E}[\mathcal{F}_{a_\triangle^*}(\overline{Y})^2]}{1 + \delta\mathbb{E}[(\mathbb{E}[\delta/\overline{\Sigma}]\overline{G}^2 - 1)\mathcal{F}_{a_\triangle^*}(\overline{Y})^2]}\right)^{1/2}.$$

Theorem F.1 (Whitened spectral estimator). *Consider the above setting and let assumptions* (A1) *to* (A5) *hold. Suppose* $a_\triangle^* > a_\triangle^\circ$. *Then, the top two eigenvalues* $\lambda_1(D)$, $\lambda_2(D)$ *of* D *satisfy*

$$\underset{d\to\infty}{\text{p-lim}}\,\lambda_1(D) = \zeta(a_\triangle^*), \quad \lim_{d\to\infty}\lambda_2(D) = \zeta(a_\triangle^\circ) \qquad \textit{almost surely,}$$

and $\zeta(a_\triangle^*) > \zeta(a_\triangle^\circ)$. *Furthermore, the limiting overlap between the spectral estimator* $\beta_\triangle^{\mathrm{spec}} = \Sigma^{-1/2}v_1(D_\triangle)$ *and* β^* *equals*

$$\underset{d\to\infty}{\text{p-lim}}\,\frac{|\langle\beta_\triangle^{\mathrm{spec}}, \beta^*\rangle|}{\|\beta_\triangle^{\mathrm{spec}}\|_2\|\beta^*\|_2} = \eta_\triangle > 0.$$

We emphasize that, even if the spectral estimator is now computed with respect to $\widetilde{X}$ whose rows have identity covariance, the observation y still depends on Σ through $y = q(\widetilde{X}\Sigma^{1/2}\beta^*, \varepsilon)$ and there is no easy way to further invert out $\Sigma^{1/2}$ therein. Thus, we cannot reduce to the $\Sigma = I_d$ case studied in [55, 62], and we follow a strategy similar to that described in Section 5 to prove Theorem 3.1.

Proof of Theorem F.1. Consider the generic GAMP iteration in (5.1). Let $\mathcal{F}_\triangle\colon \mathbb{R} \to \mathbb{R}$ be an auxiliary preprocessing function to be chosen later. Set

$$f_{t+1}(v^{t+1}) = \frac{v^{t+1}}{\beta_{t+1}}, \quad t \geq 0, \tag{F.2}$$

for a sequence $(\beta_{t+1})_{t\geq 0}$ to be specified later via state evolution. One should think of the normalization $\beta_{t+1} > 0$ as $\beta_{t+1} = \lim_{d\to\infty}\|v^{t+1}\|_2/\sqrt{d}$, so that

$$\lim_{d\to\infty}\frac{1}{\sqrt{d}}\|f_{t+1}(v^{t+1})\|_2 = 1,$$

as in (5.3). Furthermore, we set

$$g_t(u^t; y) = F_\triangle u^t, \quad t \geq 0, \tag{F.3}$$

where $F_\triangle = \mathrm{diag}(\mathcal{F}_\triangle(y)) \in \mathbb{R}^{n\times n}$ and $\mathcal{F}_\triangle(y) \in \mathbb{R}^n$ is obtained by applying $\mathcal{F}_\triangle$ to each entry of y. The coefficients b_{t+1}, c_t specialize to

$$b_{t+1} = \frac{1}{\delta\beta_{t+1}}, \quad c_t = \mathbb{E}\big[\mathcal{F}_\triangle(\bar{Y})\big] =: c.$$

Following the argument of Section 5.5, we can show that $u^t, v^{t+1}, \beta_{t+1}$ converge respectively to $u \in \mathbb{R}^n$, $v \in \mathbb{R}^d$, $\beta \in \mathbb{R}$ in the following sense

$$\lim_{t\to\infty}\lim_{n\to\infty}\frac{1}{\sqrt{n}}\|u^t - u\|_2 = 0, \quad \lim_{t\to\infty}\lim_{d\to\infty}\frac{1}{\sqrt{d}}\|v^{t+1} - v\|_2 = 0,$$

$$\lim_{t\to\infty}|\beta_{t+1} - \beta| = 0.$$

Then in the $t \to \infty$ limit, the GAMP iteration becomes

$$u = \frac{1}{\beta}\widetilde{X}v - bF_{\vartriangle}u, \quad v = \widetilde{X}^{\top}F_{\vartriangle}u - \frac{1}{\beta}cv,$$

where $b = 1/\delta\beta$ is the limit of b_{t+1} as $t \to \infty$. Solving u in terms of v from the first equation, we get

$$u = \frac{1}{\beta}(I_n + bF_{\vartriangle})^{-1}\widetilde{X}v.$$

We then use this to eliminate u from the equation for v and obtain:

$$(\beta + c)v = \widetilde{X}^{\top}F_{\vartriangle}(I_n + bF_{\vartriangle})^{-1}\widetilde{X}v. \tag{F.4}$$

Our aim is to choose $\mathcal{F}_{\vartriangle}$ judiciously to turn the above equation into an eigenequation for $D_{\vartriangle} = \widetilde{X}^{\top}T\widetilde{X}$. First, to simplify the derivation, we require $b = 1$ which will be the case if $\beta = 1/\delta$. Next, we choose

$$\mathcal{F}_{\vartriangle}(\cdot) = \mathcal{F}_{a_{\vartriangle}^*}(\cdot), \tag{F.5}$$

where the right-hand side is defined in (2.3) and $a_{\vartriangle}^*$ is to be specified later. With these choices, (F.4) becomes

$$\left(\frac{1}{\delta} + c\right)v = \frac{1}{a_{\vartriangle}^*}\widetilde{X}^{\top}T\widetilde{X}v = \frac{1}{a_{\vartriangle}^*}D_{\vartriangle}v,$$

which, upon multiplying by $a_{\vartriangle}^*$ on both sides, is an eigenequation of $D_{\vartriangle}$ with respect to the eigenvalue

$$a_{\vartriangle}^*\left(\frac{1}{\delta} + c\right) = a_{\vartriangle}^*\left(\frac{1}{\delta} + \mathbb{E}\big[\mathcal{F}_{a_{\vartriangle}^*}(\bar{Y})\big]\right),$$

and the corresponding eigenvector (up to scaling) v. The value of $a_{\vartriangle}^*$ is fixed when we enforce $\beta = 1/\delta$ which in turn enforces $b = 1$. From the state evolution analysis presented below, β can be derived and therefore $a_{\vartriangle}^*$ is defined as the solution to

$$\beta = \lim_{t\to\infty}\beta_{t+1} = \mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\bar{G}^2 - 1\right)\mathcal{F}_{a_{\vartriangle}^*}(\bar{Y})\right] = \frac{1}{\delta}. \tag{F.6}$$

Let us consider the unique solution $a_{\vartriangle}^*$ to (F.6) in $(\sup \operatorname{supp}(\mathcal{T}(\bar{Y})), \infty)$ and let $\mathcal{F}_{a_{\vartriangle}^*}: \mathbb{R} \to \mathbb{R}$ be defined in (F.5). Set the denoisers $(f_{t+1}, g_t)_{t\geq 0}$ in (5.1) to those given in (F.2) and (F.3) and initialize the GAMP iteration with

$$\widetilde{u}^{-1} = 0_n, \quad \widetilde{v}^0 = \mu\widetilde{\beta}^* + \sqrt{1 - \mu^2\mathbb{E}[\bar{\Sigma}]}w \in \mathbb{R}^d, \tag{F.7}$$

where $w \sim \mathcal{N}(0_d, I_d)$ is independent of everything else and μ is given in (F.8) below. Given all these configurations, the state evolution recursion specializes to

$$\mu_t = \frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[(\widetilde{\mathfrak{B}}^*)^{\top}\frac{V_t}{\beta_t}\right] = \frac{\delta}{\mathbb{E}[\bar{\Sigma}]}\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\big[(\widetilde{\mathfrak{B}}^*)^{\top}\widetilde{\mathfrak{B}}^*\big]\frac{\chi_t}{\beta_t} = \frac{\chi_t}{\beta_t},$$

$$\sigma_{U,t}^2 = \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left[V_t^\top \frac{V_t}{\beta_t^2}\right] - \frac{\mathbb{E}[\bar{\Sigma}]}{\delta} \mu_t^2$$

$$= \frac{1}{\beta_t^2} \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left[(\tilde{\mathfrak{B}}^*)^\top \tilde{\mathfrak{B}}^*\right] \chi_t^2 + \frac{1}{\beta_t^2} \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left[W_{V,t}^\top W_{V,t}\right] \sigma_{V,t}^2 - \frac{\mathbb{E}[\bar{\Sigma}]}{\delta} \mu_t^2$$

$$= \frac{1}{\beta_t^2} \frac{\mathbb{E}[\bar{\Sigma}]}{\delta} \chi_t^2 + \frac{1}{\beta_t^2} \frac{1}{\delta} \sigma_{V,t}^2 - \frac{\mathbb{E}[\bar{\Sigma}]}{\delta} \mu_t^2 = \frac{\sigma_{V,t}^2}{\delta \beta_t^2},$$

$$\chi_{t+1} = \frac{\delta}{\mathbb{E}[\bar{\Sigma}]} \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left[G^\top F_\vartriangle U_t\right] - \mu_t \mathbb{E}\left[\mathscr{F}_{a_\vartriangle^*}(\bar{Y})\right]$$

$$= \frac{\delta}{\mathbb{E}[\bar{\Sigma}]} \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left[G^\top F_\vartriangle G\right] \mu_t - \mu_t \mathbb{E}\left[\mathscr{F}_{a_\vartriangle^*}(\bar{Y})\right]$$

$$= \mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]} \bar{G}^2 - 1\right) \mathscr{F}_{a_\vartriangle^*}(\bar{Y})\right] \mu_t = \mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]} \bar{G}^2 - 1\right) \mathscr{F}_{a_\vartriangle^*}(\bar{Y})\right] \frac{\chi_t}{\beta_t},$$

$$\sigma_{V,t+1}^2 = \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left[U_t^\top F_\vartriangle^2 U_t\right]$$

$$= \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left[G^\top F_\vartriangle^2 G\right] \mu_t^2 + \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left[W_{U,t}^\top F_\vartriangle^2 W_{U,t}\right] \sigma_{U,t}^2$$

$$= \mathbb{E}\left[\bar{G}^2 \mathscr{F}_{a_\vartriangle^*}(\bar{Y})^2\right] \mu_t^2 + \mathbb{E}\left[\mathscr{F}_{a_\vartriangle^*}(\bar{Y})^2\right] \sigma_{U,t}^2$$

$$= \mathbb{E}\left[\bar{G}^2 \mathscr{F}_{a_\vartriangle^*}(\bar{Y})^2\right] \frac{\chi_t^2}{\beta_t^2} + \mathbb{E}\left[\mathscr{F}_{a_\vartriangle^*}(\bar{Y})^2\right] \frac{\sigma_{V,t}^2}{\delta \beta_t^2},$$

$$\beta_{t+1}^2 = \lim_{d \to \infty} \frac{1}{d} \mathbb{E}\left[V_{t+1}^\top V_{t+1}\right]$$

$$= \lim_{d \to \infty} \frac{1}{d} \mathbb{E}\left[(\tilde{\mathfrak{B}}^*)^\top \tilde{\mathfrak{B}}^*\right] \chi_{t+1}^2 + \lim_{d \to \infty} \frac{1}{d} \mathbb{E}\left[W_{V,t+1}^\top W_{V,t+1}\right] \sigma_{V,t+1}^2$$

$$= \mathbb{E}[\bar{\Sigma}] \chi_{t+1}^2 + \sigma_{V,t+1}^2.$$

There are 3 fixed points of $(\mu_t, \sigma_{U,t}, \chi_{t+1}, \sigma_{V,t+1}, \beta_{t+1})$:

$$\mathsf{FP}_+ = (\mu, \sigma_U, \chi, \sigma_V, \beta), \quad \mathsf{FP}_- = (-\mu, \sigma_U, -\chi, \sigma_V, \beta),$$

$$\mathsf{FP}_0 = \left(0, \frac{1}{\sqrt{\delta}}, 0, \frac{1}{\sqrt{\delta}} \mathbb{E}\left[\mathscr{F}_{a_\vartriangle^*}(\bar{Y})^2\right]^{1/2}, \frac{1}{\sqrt{\delta}} \mathbb{E}\left[\mathscr{F}_{a_\vartriangle^*}(\bar{Y})^2\right]^{1/2}\right),$$

where $\mu, \sigma_U, \chi, \sigma_V, \beta$ are given by

$$\beta = \mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\bar{\Sigma}]} \bar{G}^2 - 1\right) \mathscr{F}_{a_\vartriangle^*}(\bar{Y})\right] = \frac{1}{\delta},$$

$$\chi = \left(\frac{\beta^2 - \frac{1}{\delta} \mathbb{E}[\mathscr{F}_{a_\vartriangle^*}(\bar{Y})^2]}{\mathbb{E}[\bar{\Sigma}] - \frac{\mathbb{E}[\bar{\Sigma}]}{\delta \beta^2} \mathbb{E}[\mathscr{F}_{a_\vartriangle^*}(\bar{Y})^2] + \frac{1}{\beta^2} \mathbb{E}[\bar{G}^2 \mathscr{F}_{a_\vartriangle^*}(\bar{Y})^2]}\right)^{1/2}$$

$$= \left(\frac{\frac{1}{\delta^2} - \frac{1}{\delta} \mathbb{E}[\mathscr{F}_{a_\vartriangle^*}(\bar{Y})^2]}{\mathbb{E}[\bar{\Sigma}] - \delta \mathbb{E}[\bar{\Sigma}] \mathbb{E}[\mathscr{F}_{a_\vartriangle^*}(\bar{Y})^2] + \delta^2 \mathbb{E}[\bar{G}^2 \mathscr{F}_{a_\vartriangle^*}(\bar{Y})^2]}\right)^{1/2},$$

$$\sigma_V = \left(\frac{\delta^2 \mathbb{E}[\bar{G}^2 \mathscr{F}_{a_\triangle^*}(\bar{Y})^2] \chi^2}{1 - \frac{1}{\delta \beta^2} \mathbb{E}[\mathscr{F}_{a_\triangle^*}(\bar{Y})^2]} \right)^{1/2}$$

$$= \left(\frac{\mathbb{E}[\bar{G}^2 \mathscr{F}_{a_\triangle^*}(\bar{Y})^2]}{\mathbb{E}[\bar{\bar{\Sigma}}] - \delta \mathbb{E}[\bar{\bar{\Sigma}}] \mathbb{E}[\mathscr{F}_{a_\triangle^*}(\bar{Y})^2] + \delta^2 \mathbb{E}[\bar{G}^2 \mathscr{F}_{a_\triangle^*}(\bar{Y})^2]} \right)^{1/2},$$

$$\mu = \frac{\chi}{\beta} = \left(\frac{1 - \delta \mathbb{E}[\mathscr{F}_{a_\triangle^*}(\bar{Y})^2]}{\mathbb{E}[\bar{\bar{\Sigma}}] - \delta \mathbb{E}[\bar{\bar{\Sigma}}] \mathbb{E}[\mathscr{F}_{a_\triangle^*}(\bar{Y})^2] + \delta^2 \mathbb{E}[\bar{G}^2 \mathscr{F}_{a_\triangle^*}(\bar{Y})^2]} \right)^{1/2}, \qquad \text{(F.8)}$$

$$\sigma_U = \frac{\sigma_V}{\sqrt{\delta}\beta} = \left(\frac{\delta \mathbb{E}[\bar{G}^2 \mathscr{F}_{a_\triangle^*}(\bar{Y})^2]}{\mathbb{E}[\bar{\bar{\Sigma}}] - \delta \mathbb{E}[\bar{\bar{\Sigma}}] \mathbb{E}[\mathscr{F}_{a_\triangle^*}(\bar{Y})^2] + \delta^2 \mathbb{E}[\bar{G}^2 \mathscr{F}_{a_\triangle^*}(\bar{Y})^2]} \right)^{1/2}.$$

Furthermore, the initialization scheme given in (F.7) guarantees that the quintuple $(\mu_t, \sigma_{U,t}, \chi_{t+1}, \sigma_{V,t+1}, \beta_{t+1})$ stays at FP_+ for every $t \geq 0$.

Executing similar arguments in the proofs of Lemma 5.4 and of (5.34) gives

$$\lim_{t \to \infty} \operatorname*{p-lim}_{d \to \infty} \frac{\langle v^{t+1}, v_1(D_\triangle) \rangle^2}{\|v^{t+1}\|_2^2 \|v_1(D_\triangle)\|_2^2} = 1,$$

$$\operatorname*{p-lim}_{d \to \infty} \lambda_1(D_\triangle) = \zeta(a_\triangle^*) > \zeta(a_\triangle^\circ) = \lim_{d \to \infty} \lambda_2(D_\triangle). \qquad \text{(F.9)}$$

Recall from (F.1) that the whitened spectral estimator is defined as

$$\beta_\triangle^{\mathrm{spec}} = \Sigma^{-1/2} v_1(D_\triangle).$$

Given the result in (F.9), the overlap between $\beta_\triangle^{\mathrm{spec}}$ and β^* is asymptotically the same as that between $\Sigma^{-1/2} v^{t+1}$ and β^* which we compute below:

$$\lim_{t \to \infty} \operatorname*{p-lim}_{d \to \infty} \frac{\langle \Sigma^{-1/2} v^{t+1}, \beta^* \rangle^2}{\|\Sigma^{-1/2} v^{t+1}\|_2^2 \|\beta^*\|_2^2} = \frac{\lim_{t \to \infty} \operatorname*{p-lim}_{d \to \infty} \frac{1}{d^2} \langle \Sigma^{-1/2} v^{t+1}, \beta^* \rangle^2}{\lim_{t \to \infty} \operatorname*{p-lim}_{d \to \infty} \frac{1}{d} \|\Sigma^{-1/2} v^{t+1}\|_2^2},$$

the numerator and denominator of which are given respectively as follows:

$$\lim_{t \to \infty} \operatorname*{p-lim}_{d \to \infty} \frac{1}{d^2} \langle \Sigma^{-1/2} v^{t+1}, \beta^* \rangle^2 = \lim_{t \to \infty} \lim_{d \to \infty} \frac{1}{d^2} \mathbb{E}\big[(\widetilde{\mathfrak{B}}^*)^\top \Sigma^{-1/2} \mathfrak{B}^*\big]^2 \chi_{t+1}^2$$

$$= \chi^2,$$

$$\lim_{t \to \infty} \operatorname*{p-lim}_{d \to \infty} \frac{1}{d} \|\Sigma^{-1/2} v^{t+1}\|_2^2 = \lim_{t \to \infty} \lim_{d \to \infty} \frac{1}{d} \mathbb{E}\big[(\widetilde{\mathfrak{B}}^*)^\top \Sigma^{-1} \widetilde{\mathfrak{B}}^*\big] \chi_{t+1}^2$$

$$+ \frac{1}{d} \mathbb{E}\big[W_{V,t+1}^\top \Sigma^{-1} W_{V,t+1}\big] \sigma_{V,t+1}^2$$

$$= \chi^2 + \mathbb{E}\Big[\frac{1}{\bar{\bar{\Sigma}}}\Big] \sigma_V^2.$$

Using the expressions of χ, σ_V, we obtain

$$
\begin{aligned}
\operatorname*{p-lim}_{d \to \infty} \frac{\langle \beta_{\vartriangle}^{\text{spec}}, \beta^* \rangle^2}{\|\beta_{\vartriangle}^{\text{spec}}\|_2^2 \|\beta^*\|_2^2} &= \frac{\chi^2}{\chi^2 + \mathbb{E}[1/\overline{\Sigma}]\sigma_V^2} \\
&= \frac{\frac{1}{\delta^2} - \frac{1}{\delta}\mathbb{E}[\mathcal{F}_{a_{\vartriangle}^*}(\overline{Y})^2]}{\frac{1}{\delta^2} - \frac{1}{\delta}\mathbb{E}[\mathcal{F}_{a_{\vartriangle}^*}(\overline{Y})^2] + \mathbb{E}[1/\overline{\Sigma}]\mathbb{E}[\overline{G}^2 \mathcal{F}_{a_{\vartriangle}^*}(\overline{Y})^2]} \\
&= \frac{1 - \delta\mathbb{E}[\mathcal{F}_{a_{\vartriangle}^*}(\overline{Y})^2]}{1 + \delta\mathbb{E}[(\mathbb{E}[\delta/\overline{\Sigma}]\overline{G}^2 - 1)\mathcal{F}_{a_{\vartriangle}^*}(\overline{Y})^2]} = \eta_{\vartriangle},
\end{aligned}
$$

which concludes the proof. ∎

G. Auxiliary results

Proposition G.1 ($w_1 > 0$). *Let w_1 be defined in* (3.9)*. Then $w_1 > 0$.*

Proof. By definition, we have

$$
\begin{aligned}
w_1 = {}& \frac{1}{\mathbb{E}[\overline{\Sigma}]^2}\mathbb{E}\big[\overline{G}^2 \mathcal{F}_{a^*}(\overline{Y})^2\big]\mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma}}\right]^2 \\
&- \frac{1}{\delta\mathbb{E}[\overline{\Sigma}]}\mathbb{E}\big[\mathcal{F}_{a^*}(\overline{Y})^2\big]\mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma}}\right]^2 \\
&+ \frac{1}{\delta}\mathbb{E}\big[\mathcal{F}_{a^*}(\overline{Y})^2\big]\mathbb{E}\left[\frac{\overline{\Sigma}^3}{(\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma})^2}\right].
\end{aligned}
$$

The first term is strictly positive. It suffices to show that the sum of the last two terms is non-negative. This follows from the Cauchy–Schwarz inequality:

$$
\begin{aligned}
\mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma}}\right]^2 &= \mathbb{E}\left[\overline{\Sigma}^{1/2} \cdot \frac{\overline{\Sigma}^{3/2}}{\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma}}\right]^2 \\
&\leq \mathbb{E}[\overline{\Sigma}]\mathbb{E}\left[\frac{\overline{\Sigma}^3}{(\gamma^* - \mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})]\overline{\Sigma})^2}\right].
\end{aligned}
$$

Rearranging terms and noting that the common factor $\frac{1}{\delta}\mathbb{E}[\mathcal{F}_{a^*}(\overline{Y})^2]$ in the last two terms is positive, the proof is complete. ∎

Proposition G.2. *Let $W \sim P^{\otimes d}$ where P is a distribution on $\mathbb{R}$ with mean 0 and variance σ^2. Let $B \in \mathbb{R}^{d \times d}$ denote a sequence of deterministic matrices such that the empirical spectral distribution of $\frac{1}{d}B$ converges to the law of a random variable $\overline{\Sigma}$. Then*

$$
\lim_{d \to \infty} \frac{1}{d}\mathbb{E}\big[W^\top B W\big] = \sigma^2 \mathbb{E}[\overline{\Sigma}].
$$

Proof. The proof follows from a straightforward calculation:

$$\lim_{d\to\infty} \frac{1}{d} \mathbb{E}[W^\top B W] = \lim_{d\to\infty} \frac{1}{d} \sum_{i,j} \mathbb{E}[B_{i,j} W_i W_j]$$

$$= \lim_{d\to\infty} \frac{1}{d} \sum_i \mathbb{E}[W_i^2] B_{i,i} = \lim_{d\to\infty} \frac{\sigma^2}{d} \operatorname{Tr}(B) = \sigma^2 \mathbb{E}[\bar{\Sigma}]. \quad \blacksquare$$

Proposition G.3. *Let* $(G, H) \sim \mathcal{N}\left(\begin{bmatrix} 0_d \\ 0_d \end{bmatrix}, \begin{bmatrix} \sigma^2 & \rho \\ \rho & \tau^2 \end{bmatrix} \otimes I_d\right)$. *Let* $B \in \mathbb{R}^{d\times d}$ *denote a sequence of deterministic matrices such that the empirical spectral distribution of* $\frac{1}{d} B$ *converges to the law of a random variable* $\bar{\Sigma}$. *Then*

$$\lim_{d\to\infty} \frac{1}{d} \mathbb{E}[G^\top B H] = \rho \mathbb{E}[\bar{\Sigma}].$$

Proof. The proof follows from a straightforward calculation:

$$\lim_{d\to\infty} \frac{1}{d} \mathbb{E}[G^\top B H] = \lim_{d\to\infty} \frac{1}{d} \sum_{i,j} \mathbb{E}[B_{i,j} G_i H_j]$$

$$= \lim_{d\to\infty} \frac{1}{d} \sum_i \mathbb{E}[G_i H_i] B_{i,i} = \lim_{d\to\infty} \frac{\rho}{d} \operatorname{Tr}(B) = \rho \mathbb{E}[\bar{\Sigma}]. \quad \blacksquare$$

Proposition G.4 (Davis–Kahan [26]). *Let* $A, B \in \mathbb{R}^{d\times d}$ *be symmetric matrices. Then*

$$\min\{\|v_1(A) - v_1(B)\|_2, \|v_1(A) + v_1(B)\|_2\}$$
$$\leq \frac{4\|A - B\|_2}{\max\{\lambda_1(A) - \lambda_2(A), \lambda_1(B) - \lambda_2(B)\}}.$$

Note that the minimum on the left-hand side is to resolve the sign ambiguity since v is an eigenvector if and only if $-v$ is.

Remark G.1 (Spectral threshold with right rotationally invariant designs). The optimal spectral threshold for phase retrieval with right rotationally invariant designs was derived by Maillard et al. in [60, equation (11)], and this expression coincides with (3.13). To see this, note that (3.13) involves the limiting spectral distribution of Σ only through its first two moments. One can then express the same result using the limiting spectral distribution $\bar{\mu}_{X^\top X}$ of $X^\top X = \Sigma^{1/2} \tilde{X}^\top \tilde{X} \Sigma^{1/2}$, which equals the free multiplicative convolution between the Marchenko–Pastur law MP_λ (with $\lambda = 1/\delta$) and $\operatorname{law}(\bar{\Sigma})$. In particular, let $\bar{\Lambda}$ be the random variable with law $\bar{\mu}_{X^\top X}$. By using the moment-cumulant relation [71, Section 2.5] and an identity relating the square free cumulants of $\operatorname{law}(\bar{\Sigma})$ to the rectangular free cumulants of $\operatorname{law}(\bar{\Sigma}) \boxtimes \mathsf{MP}_{1/\delta}$ [7, Remark 2], we have that $\mathbb{E}[\bar{\Lambda}] = \mathbb{E}[\bar{\Sigma}]$ and $\mathbb{E}[\bar{\Lambda}^2] = \mathbb{E}[\bar{\Sigma}^2] + \frac{1}{\delta} \mathbb{E}[\bar{\Sigma}]^2$. Using these identities to write (3.13) in terms of the first two moments of $\bar{\Lambda}$, we readily obtain that this expression coincides with [60, equation (11)].

References

[1] E. Abbe, Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** (2017), article no. 177 Zbl 1403.62110 MR 3827065

[2] P. Alquier and G. Biau, Sparse single-index model. *J. Mach. Learn. Res.* **14** (2013), 243–280 Zbl 1320.62177 MR 3033331

[3] Z. Bao, Q. Han, and X. Xu, A leave-one-out approach to approximate message passing. *Ann. Appl. Probab.* **35** (2025), no. 4, 2716–2766 MR 4945090

[4] J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová, Optimal errors and phase transitions in high-dimensional generalized linear models. *Proc. Natl. Acad. Sci. USA* **116** (2019), no. 12, 5451–5460 Zbl 1416.62421 MR 3939767

[5] M. Bayati and A. Montanari, The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inform. Theory* **57** (2011), no. 2, 764–785 Zbl 1366.94079 MR 2810285

[6] S. T. Belinschi, H. Bercovici, M. Capitaine, and M. Février, Outliers in the spectrum of large deformed unitarily invariant models. *Ann. Probab.* **45** (2017), no. 6A, 3571–3625 Zbl 1412.60014 MR 3729610

[7] F. Benaych-Georges, On a surprising relation between the Marchenko–Pastur law, rectangular and square free convolutions. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** (2010), no. 3, 644–652 Zbl 1206.46055 MR 2682261

[8] E. Bolthausen, An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Comm. Math. Phys.* **325** (2014), no. 1, 333–366 Zbl 1288.82038 MR 3147441

[9] P. Boufounos and R. G. Baraniuk, 1-bit compressive sensing. In *2008 42nd Annual Conference on Information Sciences and Systems*, pp. 16–21, IEEE, 2008.

[10] Z. Bu, J. M. Klusowski, C. Rush, and W. J. Su, Algorithmic analysis and statistical estimation of SLOPE via approximate message passing. *IEEE Trans. Inform. Theory* **67** (2021), no. 1, 506–537 Zbl 1473.62238 MR 4231969

[11] Z. Bu, J. M. Klusowski, C. Rush, and W. J. Su, Characterizing the SLOPE trade-off: a variational perspective and the Donoho–Tanner limit. *Ann. Statist.* **51** (2023), no. 1, 33–61 Zbl 1539.62209 MR 4564848

[12] C. Cademartori and C. Rush, A non-asymptotic analysis of generalized vector approximate message passing algorithms with rotationally invariant designs. *IEEE Trans. Inform. Theory* **70** (2024), no. 8, 5811–5856 Zbl 7885051 MR 4773211

[13] E. J. Candès, X. Li, and M. Soltanolkotabi, Phase retrieval from coded diffraction patterns. *Appl. Comput. Harmon. Anal.* **39** (2015), no. 2, 277–299 Zbl 1329.78056 MR 3352016

[14] E. J. Candès, X. Li, and M. Soltanolkotabi, Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Trans. Inform. Theory* **61** (2015), no. 4, 1985–2007 Zbl 1359.94069 MR 3332993

[15] E. J. Candès, T. Strohmer, and V. Voroninski, PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming. *Comm. Pure Appl. Math.* **66** (2013), no. 8, 1241–1274 Zbl 1335.94013 MR 3069958

[16] M. Celentano and A. Montanari, Correlation adjusted debiased Lasso: Debiasing the Lasso with inaccurate covariate model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** (2024), no. 5, 1455–1482 MR 4896643

[17] M. Celentano, A. Montanari, and Y. Wei, The Lasso with general Gaussian designs with applications to hypothesis testing. *Ann. Statist.* **51** (2023), no. 5, 2194–2220 Zbl 1539.62222 MR 4678801

[18] M. Celentano, A. Montanari, and Y. Wu, The estimation error of general first order methods. In *Proceedings of the 33rd Conference on Learning Theory*, pp. 1078–1141, Proceedings of Machine Learning Research 125, PMLR, 2020.

[19] C. C. Chang, C. C. Chow, L. C. A. M. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* **4** (2015), no. 1

[20] S. Chen and R. Meka, Learning polynomials in few relevant dimensions. In *Proceedings of the 33rd Conference on Learning Theory*, pp. 1161–1227, Proceedings of Machine Learning Research 125, PMLR, 2020.

[21] Y. Chen and E. J. Candès, Solving random quadratic systems of equations is nearly as easy as solving linear systems. *Comm. Pure Appl. Math.* **70** (2017), no. 5, 822–883 Zbl 1379.90024 MR 3628877

[22] Y. Chen, Y. Chi, J. Fan, and C. Ma, Spectral methods for data science: A statistical perspective. *Found. Trends Mach. Learn.* **14** (2021), no. 5, 566–806, 2021 Zbl 1524.68267

[23] Y. Chen, J. Fan, C. Ma, and K. Wang, Spectral method and regularized MLE are both optimal for top-K ranking. *Ann. Statist.* **47** (2019), no. 4, 2204–2235 Zbl 1425.62038 MR 3953449

[24] R. Couillet and W. Hachem, Analysis of the limiting spectral measure of large random matrices of the separable covariance type. *Random Matrices Theory Appl.* **3** (2014), no. 4, article no. 1450016 Zbl 1305.15078 MR 3279621

[25] A. Damian, L. Pillaud-Vivien, J. Lee, and J. Bruna, Computational-statistical gaps in gaussian single-index models (extended abstract). In *Proceedings of the 37th Conference on Learning Theory*, pp. 1262–1262, Proceedings of Machine Learning Research 247, PMLR, 2024.

[26] C. Davis and W. M. Kahan, The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* **7** (1970), 1–46 Zbl 0198.47201 MR 0264450

[27] X. Ding and H. C. Ji, Spiked multiplicative random matrices and principal components. *Stochastic Process. Appl.* **163** (2023), 25–60 Zbl 1525.15027 MR 4600155

[28] X. Ding and F. Yang, Spiked separable covariance matrices and principal components. *Ann. Statist.* **49** (2021), no. 2, 1113–1138 Zbl 1469.15035 MR 4255121

[29] X. Ding and F. Yang, Tracy–Widom distribution for heterogeneous Gram matrices with applications in signal detection. *IEEE Trans. Inform. Theory* **68** (2022), no. 10, 6682–6715 Zbl 1533.94013 MR 4502560

[30] D. L. Donoho, M. Gavish, and E. Romanov, ScreeNOT: Exact MSE-optimal singular value thresholding in correlated noise. *Ann. Statist.* **51** (2023), no. 1, 122–148 Zbl 1539.62178 MR 4564851

[31] D. L. Donoho, A. Maleki, and A. Montanari, Message passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA* **106** (2009), no. 45, 18914–18919

[32] D. L. Donoho and A. Montanari, High dimensional robust M-estimation: Asymptotic variance via approximate message passing. *Probab. Theory Related Fields* **166** (2016), no. 3-4, 935–969 Zbl 1357.62220 MR 3568043

[33] R. Dudeja, M. Bakhshizadeh, J. Ma, and A. Maleki, Analysis of spectral methods for phase retrieval with random orthogonal matrices. *IEEE Trans. Inform. Theory* **66** (2020), no. 8, 5182–5203 Zbl 1446.94022 MR 4130669

[34] R. Dudeja, Y. M. Lu, and S. Sen, Universality of approximate message passing with semirandom matrices. *Ann. Probab.* **51** (2023), no. 5, 1616–1683 Zbl 1533.82001 MR 4642220

[35] R. Dudeja, J. Ma, and A. Maleki, Information theoretic limits for phase retrieval with subsampled Haar sensing matrices. *IEEE Trans. Inform. Theory* **66** (2020), no. 12, 8002–8045 Zbl 1457.94039 MR 4191039

[36] H. Eftekhari, M. Banerjee, and Y. Ritov, Inference in high-dimensional single-index models under symmetric designs. *J. Mach. Learn. Res.* **22** (2021), article no. 27 Zbl 1539.62074 MR 4253720

[37] Z. Fan, Approximate message passing algorithms for rotationally invariant matrices. *Ann. Statist.* **50** (2022), no. 1, 197–224 Zbl 1486.94026 MR 4382014

[38] Z. Fan, Y. Sun, and Z. Wang, Principal components in linear mixed models with general bulk. *Ann. Statist.* **49** (2021), no. 3, 1489–1513 Zbl 1475.62107 MR 4302572

[39] A. Fannjiang and T. Strohmer, The numerics of phase retrieval. *Acta Numer.* **29** (2020), 125–228 Zbl 07674563 MR 4189292

[40] O. Y. Feng, R. Venkataramanan, C. Rush, and R. J. Samworth, A unifying tutorial on approximate message passing. *Found. Trends Mach. Learn.* **15** (2022), no. 4, 335–536 Zbl 1491.68152

[41] R. Ganti, N. Rao, L. Balzano, R. Willett, and R. Nowak, On learning high dimensional structured single index models. *Proceedings of the AAAI Conference on Artificial Intelligence* **31** (2017), no. 1, 1898–1904

[42] C. Gerbelot and R. Berthier, Graph-based approximate message passing iterations. *Inf. Inference* **12** (2023), no. 4, 2562–2628 Zbl 07858895 MR 4644961

[43] A. Javanmard and A. Montanari, Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** (2014), 2869–2909 Zbl 1319.62145 MR 3277152

[44] A. Javanmard and A. Montanari, Hypothesis testing in high-dimensional regression under the Gaussian random design model: Asymptotic theory. *IEEE Trans. Inform. Theory* **60** (2014), no. 10, 6522–6554 Zbl 1360.62074 MR 3265038

[45] A. Javanmard and A. Montanari, Debiasing the Lasso: Optimal sample size for Gaussian designs. *Ann. Statist.* **46** (2018), no. 6A, 2593–2622 Zbl 1407.62270 MR 3851749

[46] Y. Kabashima, A CDMA multiuser detection algorithm on the basis of belief propagation. *J. Phys. A* **36** (2003), no. 43, 11111–11121 Zbl 1081.94509 MR 2025247

[47] J. Ko, F. Krzakala, and A. Pak, Optimal algorithms for the inhomogeneous Spiked Wigner Model. In *Advances in Neural Information Processing Systems 36*, pp. 76409–76424, Curran Associates, Red Hook, NY, 2023

[48] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, Probabilistic reconstruction in compressed sensing: Algorithms, phase diagrams, and threshold achieving matrices. *J. Stat. Mech.* (2012), no. 08, article no. P08009

[49] G. Li, W. Fan, and Y. Wei, Approximate message passing from random initialization with applications to $\mathbb{Z}_2$ synchronization. *Proc. Natl. Acad. Sci. USA* **120** (2023), no. 31, article no. e2302930120 Zbl 1556.62074 MR 4637851

[50] G. Li, Y. Gu, and Y. M. Lu, Phase retrieval using iterative projections: Dynamics in the large systems limit. In *53rd Annual Allerton Conference on Communication, Control, and Computing*, pp. 1114–1118, IEEE, 2015

[51] G. Li and Y. Wei, A non-asymptotic framework for approximate message passing in spiked models. [v1] 2022, [v2] 2023, arXiv:2208.03313v2

[52] Y. Li and Y. Wei, Minimum ℓ_1-norm interpolators: Precise asymptotics and multiple descent. 2021, arXiv:2110.09502v1

[53] Z. Liao and M. W. Mahoney, Hessian eigenspectra of more realistic nonlinear models. In *Advances in Neural Information Processing Systems 34*, pp. 20104–20117, Curran Associates, Red Hook, NY, 2021

[54] J. Lonsdale, J. Thomas, M. Salvatore, et al., The genotype-tissue expression (GTEx) project. *Nature Genetics* **45** (2013), no. 6, 580–585

[55] Y. M. Lu and G. Li, Phase transitions of spectral initialization for high-dimensional nonconvex estimation. *Inf. Inference* **9** (2020), no. 3, 507–541 Zbl 1470.94049 MR 4146345

[56] W. Luo, W. Alghamdi, Y. M. Lu, Optimal spectral initialization for signal recovery with applications to phase retrieval. *IEEE Trans. Signal Process.* **67** (2019), no. 9, 2347–2456 Zbl 1458.94111 MR 3962211

[57] J. Ma, R. Dudeja, J. Xu, A. Maleki, and X. Wang, Spectral method for phase retrieval: An expectation propagation perspective. *IEEE Trans. Inform. Theory* **67** (2021), no. 2, 1332–1355 Zbl 1465.94024 MR 4232014

[58] J. Ma, J. Xu, and A. Maleki, Analysis of sensing spectral for signal recovery under a generalized linear model. In *Advances in Neural Information Processing Systems 34*, pp. 22601–22613, MIT Press, Cambridge, MA, 2021

[59] A. Maillard, F. Krzakala, Y. M. Lu, and L. Zdeborová, Construction of optimal spectral methods in phase retrieval. In *Proceedings of the 2nd Mathematical and Scientific Machine Learning Conference*, pp. 693–720, Proceedings of Machine Learning Research 145, PMLR, 2022

[60] A. Maillard, B. Loureiro, F. Krzakala, and L. Zdeborová, Phase retrieval in high dimensions: Statistical and computational phase transitions. In *Advances in Neural Information Processing Systems 33*, pp. 11071–11082, Curran Associates, Red Hook, NY, 2020

[61] P. McCullagh and J. A. Nelder, *Generalized linear models*. Monogr. Statist. Appl. Probab. 37, Chapman and Hall, London, 1989 Zbl 0744.62098 MR 3223057

[62] M. Mondelli and A. Montanari, Fundamental limits of weak recovery with applications to phase retrieval. *Found. Comput. Math.* **19** (2019), no. 3, 703–773 Zbl 1464.62296 MR 3958805

[63] M. Mondelli, C. Thrampoulidis, and R. Venkataramanan, Optimal combination of linear and spectral estimators for generalized linear models. *Found. Comput. Math.* **22** (2022), no. 5, 1513–1566 Zbl 07605399 MR 4498440

[64] A. Montanari and E. Richard, A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems 27*, Curran Associates, Red Hook, NY, 2014

[65] A. Montanari and R. Venkataramanan, Estimation of low-rank matrices via approximate message passing. *Ann. Statist.* **49** (2021), no. 1, 321–345 Zbl 1461.62070 MR 4206680

[66] M. Mondelli and R. Venkataramanan, Approximate message passing with spectral initialization for generalized linear models. *J. Stat. Mech. Theory Exp.* (2022), no. 11, article no. 114003 Zbl 1539.94008 MR 4535574

[67] A. Montanari and Y. Wu, Adversarial examples in random neural networks with general activations. *Math. Stat. Learn.* **6** (2023), no. 1-2, 143–200 Zbl 1524.68323 MR 4619823

[68] A. Montanari and Y. Wu, Statistically optimal first-order algorithms: A proof via orthogonalization. *Inf. Inference* **13** (2024), no. 4, article no. iaae027 Zbl 1555.62071 MR 4806510

[69] P. Netrapalli, P. Jain, and S. Sanghavi, Phase retrieval using alternating minimization. *IEEE Trans. Signal Process.* **63** (2015), no. 18, 4814–4826 Zbl 1394.94421 MR 3385838

[70] A. Ng, M. Jordan, and Y. Weiss, On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems 14*, MIT Press, Cambridge, MA, 2001

[71] J. Novak, Three lectures on free probability. In *Random matrix theory, interacting particle systems, and integrable systems*, pp. 309–383, Math. Sci. Res. Inst. Publ. 65, Cambridge University Press, New York, 2014 Zbl 1338.46075 MR 3380692

[72] A. Pananjady and D. P. Foster, Single-index models in the high signal regime. *IEEE Trans. Inform. Theory* **67** (2021), no. 6, Part 2, 4092–4124 Zbl 1475.94043 MR 4289367

[73] D. Paul and J. W. Silverstein, No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix. *J. Multivariate Anal.* **100** (2009), no. 1, 37–57 Zbl 1154.60320 MR 2460475

[74] Y. Polyanskiy and Y. Wu, *Information theory: From coding to learning*. Cambridge, Cambridge University Press, 2024 Zbl 07887869

[75] P. Radchenko, High dimensional single index models. *J. Multivariate Anal.* **139** (2015), 266–282 Zbl 1328.62482 MR 3349492

[76] S. Rangan, Generalized approximate message passing for estimation with random linear mixing. In *2011 IEEE International Symposium on Information Theory Proceedings*, pp. 2168–2172, IEEE, 2011

[77] S. Rangan, P. Schniter, and A. K. Fletcher, Vector approximate message passing. *IEEE Trans. Inform. Theory* **65** (2019), no. 10, 6664–6684 Zbl 1432.94036 MR 4009222

[78] C. Rush, An asymptotic rate for the LASSO loss. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*, pp. 3664–3673, Proceedings of Machine Learning Research 108, PMLR, 2020.

[79] K. Sawaya, Y. Uematsu, and M. Imaizumi, High-dimensional single-index models: Link estimation and marginal inference. 2024, arXiv:2404.17812v1

[80] A. Singer, Angular synchronization by eigenvectors and semidefinite programming. *Appl. Comput. Harmon. Anal.* **30** (2011), no. 1, 20–36 Zbl 1206.90116 MR 2737931

[81] B. Su, Q. Sun, X. Yang, and B. Zhao, The exact risks of reference panel-based regularized estimators. 2024, arXiv:2401.11359v1

[82] P. Sur, A modern maximum likelihood theory for high-dimensional logistic regression. PhD thesis, Stanford University, 2019

[83] P. Sur and E. J. Candès, A modern maximum-likelihood theory for high-dimensional logistic regression. *Proc. Natl. Acad. Sci. USA* **116** (2019), no. 29, 14516–14525 Zbl 1431.62084 MR 3984492

[84] C. Thrampoulidis and A. S. Rawat, Lifting high-dimensional non-linear models with gaussian regressors. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, pp. 3206–3215, Proceedings of Machine Learning Research 89, PMLR, 2019

[85] S. van de Geer, P. Bühlmann, Y. Ritov, and R. Dezeure, On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** (2014), no. 3, 1166–1202 Zbl 1305.62259 MR 3224285

[86] R. Venkataramanan, K. Kögler, and M. Mondelli, Estimation in rotationally invariant generalized linear models via approximate message passing. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 22120–22144, Proceedings of Machine Learning Research 162, PMLR, 2022.

[87] M. J. Wainwright, Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Trans. Inform. Theory* **55** (2009), no. 5, 2183–2202 Zbl 1367.62220 MR 2729873

[88] I. Waldspurger, A. d'Aspremont, and S. Mallat, Phase recovery, MaxCut and complex semidefinite programming. *Math. Program.* **149** (2015), no. 1-2, Ser. A, 47–81 Zbl 1329.94018 MR 3300456

[89] G. Wang, G. B. Giannakis, and Y. C. Eldar, Solving systems of random quadratic equations via truncated amplitude flow. *IEEE Trans. Inform. Theory* **64** (2018), no. 2, 773–794 Zbl 1390.90451 MR 3762591

[90] T. Wang, X. Zhong, and Z. Fan, Universality of approximate message passing algorithms and tensor networks. *Ann. Appl. Probab.* **34** (2024), no. 4, 3943–3994 MR 4783034

[91] K. Wei, Solving systems of phaseless equations via Kaczmarz methods: A proof of concept study. *Inverse Problems* **31** (2015), no. 12, article no. 125008 Zbl 1332.65045 MR 3540315

[92] Y. Wu, *Information-theoretic methods for high-dimensional statistics*. Lecture Notes for ECE598YW, University of Illinois, 2016

[93] Y. Wu and K. Zhou, Lower bounds for the convergence of tensor power iteration on random overcomplete models. In *Proceedings of the 36th Conference on Learning Theory*, pp. 3783–3820, Proceedings of Machine Learning Research 195, PMLR, 2023

[94] F. Yang, Edge universality of separable covariance matrices. *Electron. J. Probab.* **24** (2019), article no. 123 Zbl 1436.15042 MR 4029426

[95] X. Yi, C. Caramanis, and S. Sanghavi, Alternating minimization for mixed linear regression. In *Proceedings of the 31st International Conference on Machine Learning*, pp. 613–621, Proceedings of Machine Learning Research 32, PMLR, 2014

[96] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah, On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** (1988), no. 4, 509–521 Zbl 0627.62022 MR 0950344

[97] C.-H. Zhang and S. S. Zhang, Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** (2014), no. 1, 217–242 Zbl 1411.62196 MR 3153940

[98] L. Zhang, Spectral analysis of large dimensional random matrices. PhD thesis, National University of Singapore, 2007

[99] Y. Zhang, M. Mondelli, and R. Venkataramanan, Precise asymptotics for spectral methods in mixed generalized linear models. [v1] 2022, [v4] 2024, arXiv:2211.11368v4

[100] Q. Zhao, P. Sur, and E. J. Candès, The asymptotic distribution of the MLE in high-dimensional logistic models: Arbitrary covariance. *Bernoulli* **28** (2022), no. 3, 1835–1861 Zbl 07526608 MR 4411513

Yihan Zhang
School of Mathematics, University of Bristol, Fry Building, Woodland Road, Bristol
BS8 1UG, UK; yihan.zhang@bristol.ac.uk

Hong Chang Ji
Department of Mathematics, University of Wisconsin-Madison, 480 Lincoln Drive, Madison,
WI 53706-1325, USA; hji56@wisc.edu

Ramji Venkataramanan
Department of Engineering, University of Cambridge, Trumpington Street, Cambridge
CB2 1PZ, UK; rv285@cam.ac.uk

Marco Mondelli
Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria;
marco.mondelli@ist.ac.at