Supplementary Material document for
Diaconis-Ylvisaker prior penalized likelihood for $p/n \to \kappa \in (0,1)$ logistic regression

Philipp Sterzinger [*1] and Ioannis Kosmidis [†2]

^{2}Department of Statistics, London School of Economics
London, WC2A 2AE, UK
^{2}Department of Statistics, University of Warwick
Coventry, CV4 7AL, UK

December 17, 2024

Contents

*p.sterzinger@lse.ac.uk
†ioannis.kosmidis@warwick.ac.uk

S1 Overview

This document contains further simulation results and the proofs of all results in the main text of *Diaconis-Ylvisaker prior penalized likelihood for $p/n \to \kappa \in (0,1)$ logistic regression*. Section S2 provides another demonstration of the signal recovery, similar to Figure 1 of the main text, simulation evidence about the stochastic boundedness of $\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\|_2/\sqrt{n}$ for various values of α following the discussion of Theorem 3.2, results on the suitability of adjusted Z-statistics for inference in support of Conjecture 6.1, and the empirical distribution of the estimator of υ from the simulation study of Section 5.1. To provide context for the proofs of our results, Appendix A restates the MDYPL setting considered in the main paper. Appendices B-F contain proofs of the main results in the paper. Key lemmas are listed in the appendices of the respective proofs and minor auxiliary lemmas are listed in Appendix G.

S2 Further numerical results and simulation experiments

S2.1 Signal recovery of the rescaled MDYPL estimator

As an additional demonstration to Figure 1, we revisit the synthetic data example of Sur and Candès (2019, Figure 2). Specifically, $\boldsymbol{x}_j \sim \mathcal{N}(\boldsymbol{0}_p, n^{-1}\boldsymbol{I}_p)$ and $\boldsymbol{\beta}_0$ consists of $n\kappa/8$ replications of $(-10, 10, 0, 0, 0, 0, 0, 0)^\top$ adequately scaled to achieve $\|\boldsymbol{\beta}_0\|_2^2 = \gamma^2$ for various values of (κ, γ) and $n = 2000$. Figure S1 shows the MDYPL estimates and their rescaled versions $\hat{\boldsymbol{\beta}}^{\mathrm{DY}}/\mu_*$ for various values of (κ, γ) and $\alpha = 1/(1+\kappa)$. The rescaled estimates are found to recover the signal $\boldsymbol{\beta}_0$ at either side of the phase transition curve as is to be expected from (x).
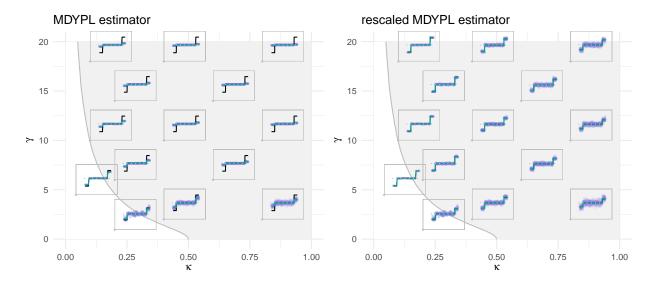
Figure S1: The MDYPL (left) and rescaled MDYPL (right) estimates for various configurations of (κ, γ) and $\alpha = 1/(1+\kappa)$, $n = 2000$ in the setting of Figure 2 in Sur and Candès (2019). Blue markers show the average coefficient estimates over 10 independent runs while the green segments denote the signal. The white and grey areas in the background indicate where the ML estimate does or does not exist asymptotically, respectively.

S2.2 Stochastic boundedness of the MDYPL estimator

To investigate the behaviour of the ℓ_2-norm of the MDYPL-estimator, we simulate $\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}(\alpha)\|_2/\sqrt{n}$, for $\alpha \in \{0.9, 0.95, 0.99, 0.99\}$ annd $\alpha = 1/(1 + \kappa)$ over a range of (κ, γ) values and $n \in \{200, 400, \ldots, 8000\}$. The covariate matrix $\boldsymbol{X}$ has i.i.d. rows $\boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{0}_p, n^{-1}\boldsymbol{I}_p)$, and the signal is the same as in Sur and Candès (2019, Figure 2) scaled to achieve the desired signal strength γ. For each specification of $\alpha, \kappa, \gamma, n$, $\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}(\alpha)\|_2/\sqrt{n}$ is estimated as the mean from 50 independent draws of $\{\boldsymbol{y}, \boldsymbol{X}\}$. The results are displayed in Figure S2 and provide no evidence that $\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}(\alpha)\|_2/\sqrt{n}$ is not bounded in probability for all (κ, γ) ranges considered, even when there is almost no shrinkage induced by the DY prior. Hence, we believe that the threshold α_0 in Theorem 3.2 is an artefact of our proof strategy.
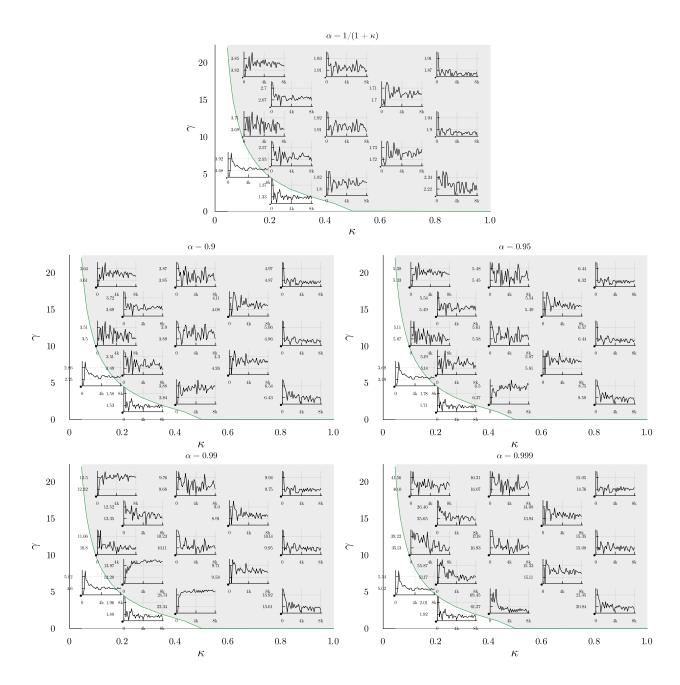
Figure S2: Stochastic boundedness of $\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\|_2/\sqrt{n}$ for various (κ, γ, α) values. Each panel shows $\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\|_2/\sqrt{n}$ for $\alpha \in \{0.9, 0.95, 0.99, (1+\kappa)^{-1}\}$ for various points in the (κ, γ) plane. In each panel, every inset plot shows the average values of $\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\left(\{1+\kappa\}^{-1}\right)\|_2/\sqrt{n}$ for $n \in \{200, 400, \ldots, 8000\}$, computed over 50 repetitions for each value of n and each (κ, γ) point, in the setting of Figure S1. The white and grey areas in the background indicate regions where the ML estimate does and does not exist asymptotically, respectively.

S2.3 Adjusted Z-statistics for MDYPL with intercept

In this section we present further simulation evidence to support Conjecture 6.1 in the main text. We fix $n = 2000$ and let p vary so that $\kappa \in \{0.2, 0.4, 0.6, 0.8\}$, setting $\alpha = (1+\kappa)^{-1}$ and consider intercept values $\theta_0 \in \{0.5, 1, \ldots, 2.5\}$. We let $\boldsymbol{X}$ be the n times p matrix with i.i.d $\mathcal{N}(0, 1/n)$ entries. $\boldsymbol{\beta}_0$ has all but $p/4$ randomly chosen entries set to zero. The nonzero entries are all of the same value and scaled such that

$\gamma^2 = \|\boldsymbol{\beta}_0\|_2^2/n = 5$. For each combination of (κ, θ_0), we draw 5000 independent copies of $\{\boldsymbol{X}, \boldsymbol{y}\}$ from a logistic regression model with intercept. We compute the adjusted Z-statistics of Theorem 3.4 with μ_*, σ_* from (15) and compare the empirical $1, 5, 10, 25, 50, 75, 90, 95, 99\%$-quantiles to those of a $\mathcal{N}(0, 1)$ distribution. Results are shown in Tables S1 and S2.

Table S1: Empirical distribution of adjusted Z-statistics from a logistic regression model with intercept using MDYPL; For $n = 2000$, $\kappa \in \{0.2, 0.4, 0.6, 0.8\}$, $\alpha = 1/(1 + \kappa)$ and $\theta_0 \in \{0.5, 1, 1.5, 2, 2.5\}$, $\boldsymbol{\beta}_0$ having $p/4$ nonzero entries of equal magnitude such that $\gamma^2 = \|\boldsymbol{\beta}_0\|_2^2/n = 5$ and $\boldsymbol{x}_j \sim \mathcal{N}(0, n^{-1}\boldsymbol{I}_p)$, we draw 5000 independent samples of $\{\boldsymbol{X}, \boldsymbol{y}\}$ according to (14) and compute the adjusted Z-statistics of Theorem 3.4 with σ_* from (15) for the first coordinate of $\boldsymbol{\beta}_0$ that is zero. For each combination of κ, β_0, we compute the average number of test statistics that lie below the $1, 5, 10, 25, 50, 75, 90, 95, 99\%$-quantiles of a $\mathcal{N}(0, 1)$ distribution.

		Percent adjusted Z-statistics $\leq \mathcal{N}(0, 1)$-quantiles								
	θ_0	1	5	10	25	50	75	90	95	99
	0.5	0.7	4.9	9.5	24.7	48.7	74.2	89.4	94.8	98.8
	1.0	1.1	5.4	10.1	24.3	49.8	74.8	89.1	94.9	99.0
$\kappa = 0.2$	1.5	1.0	5.1	10.4	25.4	50.6	75.9	90.0	94.9	99.2
	2.0	0.9	4.9	9.9	24.3	49.4	74.2	89.6	94.7	98.9
	2.5	1.3	4.8	9.8	24.9	50.2	74.8	90.2	95.2	99.2
	0.5	1.1	5.7	11.0	26.5	51.6	76.4	90.7	95.0	98.9
	1.0	1.1	5.0	9.9	24.7	49.8	76.0	90.4	95.2	99.1
$\kappa = 0.4$	1.5	0.9	4.8	10.6	25.0	50.3	75.2	89.9	94.6	98.8
	2.0	0.9	4.7	9.8	25.2	50.7	74.6	89.8	94.8	98.9
	2.5	1.1	5.3	9.7	25.0	50.1	76.0	89.9	95.3	99.1
	0.5	1.0	4.9	10.4	24.6	49.1	74.5	89.7	95.0	99.0
	1.0	1.0	5.2	10.5	25.3	49.5	75.1	90.4	94.8	99.2
$\kappa = 0.6$	1.5	1.1	5.1	10.5	25.3	49.9	75.7	90.6	95.6	99.0
	2.0	0.8	4.9	10.4	25.1	49.9	75.2	90.5	95.3	98.7
	2.5	0.9	5.0	10.1	25.3	50.3	74.9	90.2	95.1	98.9
	0.5	1.0	5.2	9.9	25.4	49.0	74.5	89.6	94.9	99.0
	1.0	1.0	4.9	10.2	25.2	49.6	74.2	90.0	95.1	98.9
$\kappa = 0.8$	1.5	0.9	4.9	10.4	26.2	50.7	75.1	89.8	95.3	98.8
	2.0	1.3	5.1	9.5	24.7	50.1	75.6	90.0	94.9	99.0
	2.5	1.2	5.3	10.4	25.0	49.5	75.3	90.9	95.6	99.1

Table S2: Empirical distribution of adjusted Z-statistics from a logistic regression model with intercept using MDYPL; For $n = 2000$, $\kappa \in \{0.2, 0.4, 0.6, 0.8\}$, $\alpha = 1/(1 + \kappa)$ and $\theta_0 \in \{0.5, 1, 1.5, 2, 2.5\}$, $\boldsymbol{\beta}_0$ having $p/4$ nonzero entries of equal magnitude such that $\gamma^2 = \|\boldsymbol{\beta}_0\|_2^2/n = 5$ and $\boldsymbol{x}_j \sim \mathcal{N}(0, n^{-1}\boldsymbol{I}_p)$, we draw 5000 independent samples of $\{\boldsymbol{X}, \boldsymbol{y}\}$ according to (14) and compute the adjusted Z-statistics of Theorem 3.4 with σ_* from (15) for the first coordinate of $\boldsymbol{\beta}_0$ that is nonzero. For each combination of κ, β_0, we compute the average number of test statistics that lie below the $1, 5, 10, 25, 50, 75, 90, 95, 99\%$-quantiles of a $\mathcal{N}(0, 1)$ distribution.

| | θ_0 | Percent adjusted Z-statistics $\leq \mathcal{N}(0,1)$-quantiles | | | | | | | | |
		1	5	10	25	50	75	90	95	99
$\kappa = 0.2$	0.5	0.7	4.7	9.2	24.5	48.8	74.4	90.0	95.2	99.1
	1.0	1.2	5.3	10.1	25.6	50.7	75.5	90.3	94.7	99.0
	1.5	1.0	5.0	10.3	24.9	49.4	73.7	89.9	94.9	98.7
	2.0	1.0	5.0	9.9	24.8	50.8	74.7	90.0	95.0	98.8
	2.5	1.0	4.7	9.5	23.8	49.1	74.7	89.7	95.4	99.1
$\kappa = 0.4$	0.5	0.9	5.3	10.2	25.6	50.5	76.3	91.0	95.5	99.2
	1.0	1.0	5.1	10.2	25.8	50.8	74.9	89.9	95.2	98.9
	1.5	1.0	5.1	10.8	25.6	49.8	75.2	89.9	95.0	99.1
	2.0	1.0	4.9	9.8	25.3	49.7	74.3	89.7	95.0	99.0
	2.5	0.8	4.5	9.6	25.1	50.9	75.9	89.7	94.8	98.9
$\kappa = 0.6$	0.5	0.9	5.3	10.2	25.7	50.9	75.8	90.3	95.0	98.9
	1.0	1.0	5.1	9.9	25.1	51.0	75.6	90.7	95.2	99.0
	1.5	0.8	5.2	10.2	25.0	50.0	75.0	89.9	95.3	98.8
	2.0	1.0	4.6	9.2	24.9	50.0	74.5	90.0	95.4	99.1
	2.5	0.9	5.1	9.8	25.0	50.7	74.8	90.2	95.2	99.2
$\kappa = 0.8$	0.5	1.2	4.8	9.2	23.9	48.4	74.0	89.7	95.1	99.1
	1.0	1.1	5.5	10.5	25.1	50.2	75.0	89.9	94.8	99.0
	1.5	1.1	5.2	9.7	24.8	50.0	75.8	90.6	95.2	99.1
	2.0	1.0	4.8	10.0	25.5	50.6	74.7	89.6	95.0	98.9
	2.5	0.8	4.3	9.6	25.2	49.9	74.9	89.8	94.7	99.2

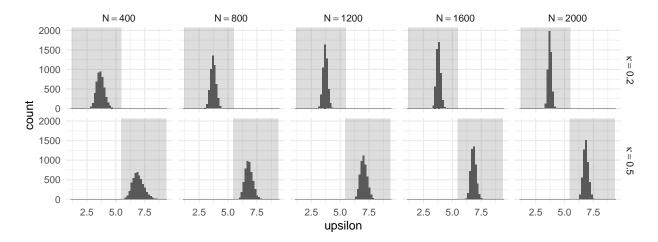## S2.4   Estimates of υ for the simulation study of Section 5.1

Figure S3: Empirical distribution of the SLOE estimator of υ in the experiments of Section 5.1. The shaded area is the image of the function $\upsilon = h(\gamma)$ for $\kappa \in \{0.2, 0.5\}$ and $\alpha = 0.95$. See Figure 8 of the main text for $\upsilon = h(\gamma)$.

A   Maximum Diaconis-Ylvisaker prior penalized likelihood

Consider data $\{\boldsymbol{y}, \boldsymbol{X}\}$ with responses $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$, $y_j \in \{0, 1\}$, and covariate matrix $\boldsymbol{X} \in \Re^{n \times p}$ with rows $\boldsymbol{x}_j \in \Re^p$ $(j = 1, \ldots, n)$. A logistic regression model assumes that conditional on $\boldsymbol{x}_j$, the responses y_j are independent realizations of Bernoulli random variables with mean $\zeta'(\boldsymbol{x}_j^\top \boldsymbol{\beta}_0)$ where $\zeta(x) = \log(1 + e^x)$ and $\boldsymbol{\beta}_0 \in \Re^p$ is the unknown p-dimensional vector of regression coefficients.

Given hyperparameters $\alpha \in (0, 1]$ and $\boldsymbol{\beta}_P \in \Re^p$, the logarithm of the Diaconis and Ylvisaker (1979) prior, bar the normalizing constant, for the logistic regression model is given by

$$\log p(\boldsymbol{\beta}; \boldsymbol{X}) = \frac{1 - \alpha}{\alpha} \sum_{j=1}^n \left\{ \zeta'\left(\boldsymbol{x}_j^\top \boldsymbol{\beta}_P\right) \boldsymbol{x}_j^\top \boldsymbol{\beta} - \zeta\left(\boldsymbol{x}_j^\top \boldsymbol{\beta}\right) \right\}, \tag{S1}$$

where $\boldsymbol{\beta}_P$ is the mode of the prior, while α controls the variability of the prior distribution. The DY-prior penalized log-likelihood, without the normalizing constant is given by

$$\ell^*(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{X}) = \frac{1}{\alpha} \sum_{j=1}^n \left\{ \left(\alpha y_j + (1 - \alpha)\zeta'\left(\boldsymbol{x}_j^\top \boldsymbol{\beta}_P\right)\right) \boldsymbol{x}_j^\top \boldsymbol{\beta} - \zeta\left(\boldsymbol{x}_j^\top \boldsymbol{\beta}\right) \right\}. \tag{S2}$$

The maximiser of (S2) is denoted by $\hat{\boldsymbol{\beta}}^{\mathrm{DY}}$. Hence, $\hat{\boldsymbol{\beta}}^{\mathrm{DY}}$ is the maximiser of a logistic regression model log-likelihood with pseudo-responses $\boldsymbol{y}^* = (y_1^*, \ldots, y_n^*)^\top$, $y_j^* = \alpha y_j + (1 - \alpha)\zeta'(\boldsymbol{x}_j^\top \boldsymbol{\beta}_P)$, $\alpha \in (0, 1]$. For the remainder of this work, $\boldsymbol{\beta}_P = \boldsymbol{0}_p$ and $\hat{\boldsymbol{\beta}}^{\mathrm{DY}}$ is understood as the maximiser of $\ell(\boldsymbol{\beta}; \boldsymbol{y}^*, \boldsymbol{X})$.

B   Proof of Theorem 3.1

As in Sur and Candès (2019), we assume that the model matrix $\boldsymbol{X}$ has independent rows $\boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{0}_p, n^{-1}\boldsymbol{I}_p)$, where $\boldsymbol{I}_p$ is the $p \times p$ identity matrix, and that the linear predictors $\boldsymbol{x}_i^\top \boldsymbol{\beta}_0$ have asymptotic variance $\lim_{n \to \infty} \mathrm{var}(\boldsymbol{x}_i^\top \boldsymbol{\beta}_0) = \lim_{n \to \infty} \|\boldsymbol{\beta}_0\|_2^2 / n = \gamma^2$. It is further assumed that there exists a random variable $\bar{\beta} \sim \pi_{\bar{\beta}}$ with finite second moment, such that the empirical distribution of the entries of $\boldsymbol{\beta}_0$ converges weakly in distribution to $\pi_{\bar{\beta}}$ and $\lim_{n \to \infty} \frac{1}{p} \sum_{j=1}^n \boldsymbol{\beta}_{0,j}^2 = \mathrm{E}(\bar{\beta}^2)$. Finally, let $\lim_{n \to \infty} p/n = \kappa$ for $\kappa \in (0, 1)$.

B.1 Asymptotic characterisation via stationary state evolution

Recall the state evolution equations (S3), (S4), (S5) below.

$$\mathrm{E}\left[2\zeta'(Z)Z\left\{\frac{1+\alpha}{2}-\zeta'\left(\mathrm{prox}_{b\zeta}\left(Z_*+\frac{1+\alpha}{2}b\right)\right)\right\}\right]=0 \tag{S3}$$

$$1-\kappa-\mathrm{E}\left[\frac{2\zeta'(Z)}{1+b\zeta''(\mathrm{prox}_{b\zeta}\left(Z_*+\frac{1+\alpha}{2}b\right))}\right]=0 \tag{S4}$$

$$\sigma^2-\frac{b^2}{\kappa^2}\mathrm{E}\left[2\zeta'(Z)\left\{\frac{1+\alpha}{2}-\zeta'\left(\mathrm{prox}_{b\zeta}\left(Z_*+\frac{1+\alpha}{2}b\right)\right)\right\}^2\right]=0\,. \tag{S5}$$

Here, $Z\sim N(0,\gamma^2)$, $Z_*=\mu Z+\kappa^{1/2}\sigma G$ for $G\sim N(0,1)$ which is independent of Z, $\zeta'(\cdot),\zeta''(\cdot)$ are the first and second derivatives of $\zeta(\cdot)$ and $\mathrm{prox}_{b_*\zeta}(x)=\arg\min_u\left\{b_*\zeta(u)+\frac{1}{2}(x-u)^2\right\}$ denotes the proximal operator. Equations (S3)-(S5) define the stationary points to the state evolution iterates (S16)-(S18) of the AMP recursion (S12)-(S13). A discussion about how these equations are derived is thus postponed to B.4.2.

Denote by $\boldsymbol{J}(\mu,b,\sigma)$ the Jacobian matrix of the LHS of (S3)-(S5) with respect to (μ,b,σ).

Theorem 3.1. *Assume that (α,κ,γ) are such that $\|\hat{\boldsymbol{\beta}}^{DY}\|_2=\mathcal{O}(n^{1/2})$ almost surely and that (S3)-(S5) admit a solution (μ_*,b_*,σ_*) such that $\boldsymbol{J}(\mu_*,b_*,\sigma_*)$ is nonsingular. Then for any function $\psi:\Re^2\to\Re$ that is pseudo-Lipschitz of order 2,*

$$\frac{1}{p}\sum_{j=1}^n\psi(\hat{\boldsymbol{\beta}}_j^{DY}-\mu_*\boldsymbol{\beta}_{0,j},\boldsymbol{\beta}_{0,j})\xrightarrow{a.s.}E\left[\psi(\sigma_*G,\bar{\beta})\right],\quad as\ n\to\infty\,, \tag{S6}$$

where $G\sim N(0,1)$ is independent of $\bar{\beta}\sim\pi_{\bar{\beta}}$.

The proof of Theorem 3.1 closely follows the proof of Theorem 1 in Sur and Candès (2019). We replace the logistic indicator function $h(u,v)=\{1-\mathrm{sign}(v\le\zeta'(u))\}/2$ is simply replaced by $h(u,v)=\alpha\{1-\mathrm{sign}(v\le\zeta'(u))\}/2+(1-\alpha)/2$ and the derivations are adjusted accordingly. Further, we provide a smoothing argument that formally justifies the asymptotic results of the relevant AMP recursion, and which has been missing hitherto.

B.2 Proof outline for Theorem 3.1

The following steps constitute the proof of Theorem 3.1:

A.2.1) *Target AMP recursion*: As a starting point, we formulate an AMP recursion, whose stationary point coincides with $\hat{\boldsymbol{\beta}}^{\mathrm{DY}}$.

A.2.2) *Centred target AMP recursion*: Based on the target AMP recursion, we devise a centred AMP recursion, for which the aggregate asymptotic behaviour can be established via existing AMP results.

A.2.3) *Matrix-valued AMP embedding*: The centred AMP recursion is embedded into a symmetric, matrix valued AMP recursion of Javanmard and Montanari (2013) for which the topical AMP asymptotics hold. A reduction argument, which is also due to Javanmard and Montanari (2013) and which was subsequently used in Donoho and Montanari (2016) and Sur and Candès (2019), identifies the correct asymptotic behaviour for the centred AMP recursion.

A.2.4) *Smoothing argument*: As the logistic indicator function is not Lipschitz, which is a prerequisite for the application of the asymptotic result of Javanmard and Montanari 2013, Theorem 1, we instead establish a Lipschitz approximation to the indicator function that is parametrised by $\epsilon>0$, such that $\epsilon=0$ recovers the original recursions of A.2.1 and A.2.2.

A.2.5) *Asymptotic equivalence of target and centred AMP*: Next, we show that the target recursion and the centred AMP recursion, whose asymptotic behaviour is characterised through A.2.3, are asymptotically equivalent for sufficiently large iterations.

A.2.6) *Asymptotic equivalence of AMP and MDYPL*: Finally, we shown that for $\epsilon \to 0$, the approximate AMP iterates are arbitrarily close to $\hat{\boldsymbol{\beta}}^{\mathrm{DY}}$, asymptotically, if iterated for sufficiently long.

B.3 Matrix-valued AMP

For the readers' convenience, this subsection introduces the matrix valued AMP recursion of Javanmard and Montanari (2013), which is the theoretical workhorse to derive the asymptotic behaviour of the AMP recursion of interest. It is more or less a verbatim replication of Javanmard and Montanari 2013, Section 2, tailored to the specific requirements of the problem at hand.

Given integers q, N, let $\mathcal{V}_{q,N} \equiv (\Re^q)^N$, so that a vector $x \in \mathcal{V}_{q,N}$ is regarded as a N-vector with entries in $\Re^q$, i.e. $x = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ where $\mathbf{x}_i \in \Re^q$ with components denoted by $\mathbf{x}_i = (\mathbf{x}_i(1), \ldots, \mathbf{x}_i(q))$ $(i = 1, \ldots, N)$. Given $x \in \mathcal{V}_{q,N}$, define its ℓ_2-norm as $\|x\|_2 = \left(\sum_{i=1}^N \|\mathbf{x}_i\|_2^2 \right)^{1/2}$. Given a matrix $\boldsymbol{A} \in \Re^{N \times N}$, let it act on $\mathcal{V}_{q,N}$ so that for $v, v' \in \mathcal{V}_{q,N}$, $v' = \boldsymbol{A}v$ for $\mathbf{v}_i' = \sum_{j=1}^N A_{ij} \mathbf{v}_j$ for $i \in [N]$, where $[N]$ denotes the integers $1, \ldots, N$. That is, $\boldsymbol{A}$ is identified with the Kronecker product $\boldsymbol{A} \otimes \boldsymbol{I}_q$ when acting on elements of $\mathcal{V}_{N,q}$.

Definition 1 (Symmetric AMP)**.** A symmetric AMP instance is a triple $(\boldsymbol{A}, \mathcal{F}, x^0)$, where

1) $\boldsymbol{A} = \boldsymbol{G} + \boldsymbol{G}^\top$, where $\boldsymbol{G} \in \Re^{N \times N}$ has i.i.d. $N(0, (2N)^{-1})$ entries

2) $\mathcal{F} = \{f^k : k \in [N]\}$ is a collection of mappings $f^k : \Re^q \times \mathbb{N} \to \Re^q$, $(\mathbf{x}, t) \mapsto f^k(\mathbf{x}, t)$ is locally Lipschitz in its first argument

3) $x^0 \in \mathcal{V}_{q,N}$ is an initial condition.

Given $\mathcal{F} = \{f^k : k \in [N]\}$, denote by $f(\cdot, t)$ the mapping from $\mathcal{V}_{q,N} \to \mathcal{V}_{q,N}$ such that $v' = f(v, t)$ for $\mathbf{v}_i' = f^i(\mathbf{v}_i, t)$ and all $i \in [N]$.

Definition 2 (AMP orbit)**.** The AMP orbit corresponding to an AMP instance $(\boldsymbol{A}, \mathcal{F}, x^0)$ is the sequence of vectors $\{x^t\}_{t \in \mathbb{N}_0}$, $x^t \in \mathcal{V}_{q,N}$ such that for $t \geq 0$,

$$x^{t+1} = \boldsymbol{A}f(x^t, t) - \boldsymbol{B}_t f(x^{t-1}, t-1), \tag{S7}$$

where $\boldsymbol{B}_t : \mathcal{V}_{q,N} \to \mathcal{V}_{q,N}$ is the linear operator such that for $v' = \boldsymbol{B}_t v$,

$$\mathbf{v}_i' = \left(\frac{1}{N} \sum_{j=1}^N \frac{\partial f^j}{\partial \mathbf{x}}(\mathbf{x}_j^t, t) \right) \mathbf{v}_i,$$

and where $\frac{\partial f^j}{\partial \mathbf{x}}$ denotes the Jacobian matrix of $f^j(\cdot, t) : \Re^q \to \Re^q$ given t.

Any expression with negative t-index is considered zero in the above definition.

Definition 3 (Converging AMP)**.** A sequence of AMP instances $\left\{ (\boldsymbol{A}(N), \mathcal{F}_N, x^{0,N}) \right\}_{N \in \mathbb{N}_0}$ is said to be converging if

i) There exist integers q, q', and

ii) a function $g : \Re^q \times \Re^q \times [q'] \times \mathbb{N} \to \Re^q$ with $g(\mathbf{x}, \mathbf{y}, a, t) = (g_1(\mathbf{x}, \mathbf{y}, a, t), \ldots, g_q(\mathbf{x}, \mathbf{y}, a, t))$, such that for each $r \in [q], a \in [q'], t \in \mathbb{N}$, $g_r(\cdot, \cdot, a, t)$ is Lipschitz,

iii) q' probability measures $P_1, \ldots, P_{q'}$ on $\Re^q$,

---

9

iv) for each N, a finite partition $C_1^N \cup \ldots \cup C_{q'}^N = [N]$,

v) q' positive definite matrices $\widehat{\boldsymbol{\Sigma}}_1^0 \ldots, \widehat{\boldsymbol{\Sigma}}_{q'}^0 \in \Re^{q \times q}$,

such that the following happens:

1) For each $a \in [q']$, one has $\lim\limits_{N \to \infty} \frac{|C_a^N|}{N} = c_a \in (0, 1)$,

2) There exists $y \in \mathcal{V}_{q,N}$ with $y = (\mathbf{y}_1, \ldots, \mathbf{y}_N), \mathbf{y}_i \in \Re^q$ $(i = 1, \ldots, N)$, such that for each $N \in \mathbb{N}_0$, $a \in [q']$, $i \in C_a^N$, $f^i(\mathbf{x}, t) = g(\mathbf{x}, \mathbf{y}_i, a, t)$. Further, the empirical distribution of $\{\mathbf{y}_i\}_{i \in C_a^N}$, denote it by $\widehat{P}_a$, converges weakly to P_a, i.e.

$$\frac{1}{|C_a^N|} \sum_{i \in C_a^N} \delta_{\mathbf{y}_i} \xrightarrow{d} P_a \,,$$

3) For each $a \in [q']$, in probability,

$$\lim_{N \to \infty} \frac{1}{|C_a^N|} \sum_{i \in C_a^N} g(\mathbf{x}_i^0, \mathbf{y}_i, a, 0) g(\mathbf{x}_i^0, \mathbf{y}_i, a, 0)^\top = \widehat{\boldsymbol{\Sigma}}_a^0 \,.$$

This definition follows Remark 1 of Javanmard and Montanari (2013) and allows for a partition size q' that is distinct from q. Further note that the functions $f^i(\cdot, \cdot)$ of Definition 3, 2), depends on $\mathbf{y}_i$, but as these do not change between iterations, this dependence is suppressed for notational convenience.

Definition 4 (State evolution). Given $q, q' \in \mathbb{N}_0$ and starting points $\{\widehat{\boldsymbol{\Sigma}}_a^0\}_{a \in [q']}, \widehat{\boldsymbol{\Sigma}}_a^0 \in \Re^{q \times q}$ positive definite, and weights $\{c_a\}_{a \in [q']}, c_a \in \Re$, the state evolution defines a sequence of positive definite matrices $\{\boldsymbol{\Sigma}^t\}_{t \in \mathbb{N}}$, such that for each $t \in \mathbb{N}$,

$$\boldsymbol{\Sigma}^t = \sum_{b=1}^{q'} c_b \widehat{\boldsymbol{\Sigma}}_b^{t-1} \tag{S8}$$

$$\widehat{\boldsymbol{\Sigma}}_a^t = \mathrm{E}\left[g(Z_a^t, Y_a, a, t) g(Z_a^t, Y_a, a, t)^\top\right] \quad (a = 1, \ldots, q'), \tag{S9}$$

where $Y_a \sim P_a$, $Z_a^t \sim N(\mathbf{0}_q, \boldsymbol{\Sigma}^t)$ and where $Y_a \perp Z_a^t$ for $a \in [q']$.

It is now possible to state Theorem 1 of Javanmard and Montanari (2013) which will serve as the blackbox result to establish the asymptotic behaviour of the MDYPL estimator. This result makes use of pseudo-Lipschitz functions, a definition of which is given for completeness.

Definition 5 (Pseudo-Lipschitzianity). For $k \geq 1$, a function $\psi : \Re^m \to \Re$ is said to be pseudo-Lipschitz of order k, denote the corresponding class of functions by $\mathrm{PL}(k)$, if there exists a constant $L > 0$ such that for all $\boldsymbol{u}, \boldsymbol{v} \in \Re^m$,

$$|\psi(\boldsymbol{u}) - \psi(\boldsymbol{v})| \leq L \left(1 + \|\boldsymbol{u}\|_2^{k-1} + \|\boldsymbol{v}\|_2^{k-1}\right) \|\boldsymbol{u} - \boldsymbol{v}\|_2 \,.$$

Hence, if $\psi \in \mathrm{PL}(k)$, there exists a constant L' such that for all $\boldsymbol{u} \in \Re^m$, $|\psi(\boldsymbol{u})| \leq L'(1 + \|\boldsymbol{u}\|_2^k)$. Further, note that every Lipschitz function is also pseudo Lipschitz for any $k \geq 1$.

Theorem B.1 (Javanmard and Montanari (2013) Theorem 1). *Let $\left\{(\boldsymbol{A}(N), \mathcal{F}_N, x^{0,N})\right\}_{N \in \mathbb{N}_0}$ be a converging sequence of AMP instances and denote by $\{x^t\}_{t \in \mathbb{N}_0}$ the corresponding AMP orbit. Suppose further that for some $k \geq 2$ and all $a \in [q']$, $E_{P_a}[\|Y_a\|^{2k-2}]$ is bounded and that $E_{\widehat{P}_a}[\|Y_a\|^{2k-2}] \to E_{P_a}[\|Y_a\|^{2k-2}]$*

as $N \to \infty$. Then for all $t \in \mathbb{N}$, $a \in [q']$ and any pseudo-Lipschitz function $\psi : \Re^q \times \Re^q \to \Re$ of order k,

$$\lim_{N \to \infty} \frac{1}{|C_a^N|} \sum_{i \in C_a^N} \psi(\mathbf{x}_i^t, \mathbf{y}_i) = E[\psi(Z_a^t, Y_a)] \,,$$

where $Y_a \sim P_a$, $Z_a^t \sim N(\mathbf{0}_q, \mathbf{\Sigma}^t)$ and $Y_a \perp Z_a^t$.

B.4 Proof of Theorem 3.1

B.4.1 Target AMP recursion

Consider the following recursion. Given initial conditions $\hat{\boldsymbol{\beta}}^0$, $\boldsymbol{S}^0 = \boldsymbol{X}\hat{\boldsymbol{\beta}}^0$, and a sequence of constants $\{\mu_t, b_t, \sigma_t\}_{t \in \mathbb{N}_0}$, for $t \in \mathbb{N}$, set

$$\hat{\boldsymbol{\beta}}^t = \hat{\boldsymbol{\beta}}^{t-1} + \kappa^{-1}\boldsymbol{X}^\top \Psi^{t-1}(\boldsymbol{y}^*, \boldsymbol{S}^{t-1}) \tag{S10}$$

$$\boldsymbol{S}^t = \boldsymbol{X}\hat{\boldsymbol{\beta}}^t - \Psi^{t-1}(\boldsymbol{y}^*, \boldsymbol{S}^{t-1}) \,, \tag{S11}$$

where $\Psi^t(\boldsymbol{y}^*, \boldsymbol{S}^t)$ acts elementwise on $\boldsymbol{y}^*, \boldsymbol{S}^t$ as

$$\Psi^t(y_i^*, \boldsymbol{S}_i^t) = b_t \left\{ y_i^* - \zeta'\left(\text{prox}_{b_t\zeta}\left(b_t y_i^* + \boldsymbol{S}_i^t\right)\right) \right\} \,,$$

for $i = 1, \ldots, n$ and where $y_i^* = \alpha y_i + (1 - \alpha)/2$. For $\alpha = 1$, this recursion corresponds to equations (49) and (50) of the supplementary material of Sur and Candès (2019).

The relation of this recursion to the maximiser of $\ell(\boldsymbol{\beta}; \boldsymbol{y}^*, \boldsymbol{X})$ is shown in Sur and Candès 2019, Appendix B: If there exists a fixed points $\hat{\boldsymbol{\beta}}_*, \boldsymbol{S}_*$ along with (μ_*, b_*, σ_*), then by construction

$$\boldsymbol{X}^\top b_* \left(\boldsymbol{y}^* - \zeta'(\text{prox}_{b_*\zeta}\left(b_* \boldsymbol{y}^* + \boldsymbol{S}_*\right)) \right) = \mathbf{0}_p$$

$$(b_* \boldsymbol{y}^* + \boldsymbol{S}_*) - b_* \zeta'(\text{prox}_{b_*\zeta}\left(b_* \boldsymbol{y}^* + \boldsymbol{S}_*\right)) = \boldsymbol{X}\hat{\boldsymbol{\beta}}_* \,.$$

Now using that $x - b\zeta'(\text{prox}_{b\zeta}(x)) = \text{prox}_{b\zeta}(x)$, it follows immediately that

$$\boldsymbol{X}^\top \left(\boldsymbol{y}^* - \zeta'(\boldsymbol{X}\hat{\boldsymbol{\beta}}_*) \right) = \mathbf{0}_p \,,$$

the left hand side of which is proportional to the gradient of (S2), and by strict concavity thereof, $\hat{\boldsymbol{\beta}}_*$ is equal to $\hat{\boldsymbol{\beta}}^{\text{DY}}$.

B.4.2 Centred target AMP recursion

Instead of analysing the target AMP recursion directly, a centred variant thereof is analysed. This centring ensures that the iterates of the resulting recursion have well-defined limits that are understood by AMP theory (See remarks in Donoho and Montanari 2016, Appendix B). In particular, as in Sur and Candès (2019) and Donoho and Montanari (2016), consider the following centred recursion with adjusted initial conditions $\boldsymbol{\nu}^0 = \hat{\boldsymbol{\beta}}^0 - \mu_0 \boldsymbol{\beta}_0$, $\boldsymbol{R}^0 = \boldsymbol{S}^0$ for $\hat{\boldsymbol{\beta}}^0, \boldsymbol{\beta}_0 \in \Re^p$, $\mu_0 \in \Re$ and $\boldsymbol{S}^0 \in \Re^n$, for $t \in \mathbb{N}$, let

$$\boldsymbol{\nu}^t = q_{t-1}(\boldsymbol{\nu}^{t-1} + \mu_{t-1}\boldsymbol{\beta}_0) - a_t\boldsymbol{\beta}_0 + \kappa^{-1}\boldsymbol{X}^\top \Psi^{t-1}(\boldsymbol{y}^*, \boldsymbol{R}^{t-1}) \tag{S12}$$

$$\boldsymbol{R}^t = \boldsymbol{X}(\boldsymbol{\nu}^t + \mu_t\boldsymbol{\beta}_0) - \Psi^{t-1}(\boldsymbol{y}^*, \boldsymbol{R}^{t-1}) \,, \tag{S13}$$

where $a_0 = \mu_0$ and for $t \in \mathbb{N}$,

$$q_t = -\frac{1}{\kappa n} \sum_{i=1}^n \frac{\partial \Psi^t}{\partial R}(y_i^*, R)\Big|_{R=\boldsymbol{R}_i^t} \tag{S14}$$

$$a_t = \frac{1}{\kappa n} \sum_{i=1}^n \frac{\partial \Psi^{t-1}}{\partial Z}(h(Z, \bar{\varepsilon}_i), \boldsymbol{R}_i^{t-1})\Big|_{Z=\boldsymbol{x}_i^\top \boldsymbol{\beta}_0} \,, \tag{S15}$$

11

where $\bar{\varepsilon}_i \sim U(0,1) \, (i = 1, \ldots, n)$ are i.i.d. draws independent of all other random variables.

Given initial conditions (μ_0, b_0, σ_0), let the state evolution of (μ_t, b_t, σ_t) be characterised by the solution to the following system of nonlinear equations

$$1 - \kappa = \mathrm{E} \left[\frac{1}{1 + b_t \zeta'' \left(\mathrm{prox}_{b_t \zeta} \left(b_t Y^* + Z_t \right) \right)} \right] \tag{S16}$$

$$\mu_{t+1} = \frac{b_t}{\kappa \gamma^2} \mathrm{E} \left[Z \left\{ Y^* - \zeta' \left(\mathrm{prox}_{b_t \zeta} \left(b_t Y^* + Z_t \right) \right) \right\} \right] + \mu_t \tag{S17}$$

$$\sigma_{t+1}^2 = \frac{b_t^2}{\kappa^2} \mathrm{E} \left[\left\{ Y^* - \zeta' \left(\mathrm{prox}_{b_t \zeta} \left(b_t Y^* + Z_t \right) \right) \right\}^2 \right] , \tag{S18}$$

for independent $Z \sim N(0, \gamma^2)$, $G \sim N(0,1)$, and $\bar{\varepsilon} \sim U(0,1)$, and letting

$$Y^* = \frac{\alpha}{2} \left\{ 1 - \mathrm{sign} \left(\bar{\varepsilon} - \zeta'(Z) \right) \right\} + \frac{1 - \alpha}{2}, \quad Z_t = \mu_t Z + \kappa^{1/2} \sigma_t G \,.$$

The iterate of σ_{t+1} in (S18) arises from the state evolution $\boldsymbol{\Sigma}^t$ (cf. Definition 4) that characterises the variance of the AMP iterates (S10) and (S11) as is shown in (S50). On the other hand, b_t, μ_t come from the nonlinear corrections a_t, q_t in (S14) and (S14) to the recursion (S10) and (S11). (S53) and (S55)-(S59), establish the relation between b_t, μ_t and the corrections a_t, q_t. These corrections ensure that the AMP iterates have a tractable asymptotic distribution. It is customary in the AMP literature to refer to the iterates μ_t, b_t, σ_t as state evolutions and equations (S36)-(S38) as state evolution equations – a convention which is henceforth adhered to in this text. Feng *et al.* (2022), Section 2.1, and Section 4 provide a thorough discussion of the state evolutions and their role in AMP.

The relation to equations (22)-(24) of the supplementary material of Sur and Candès (2019) and equation (118) in Feng *et al.* (2022) is evident.

B.4.3 Matrix-valued AMP-embedding

This section embeds the centred AMP recursion of (B.4.2) in a symmetric AMP-instance, which is shown to satisfy all properties of Definitions 1-4, apart from a Lipschitzianity constraint on $g(\cdot, \cdot, a, t)$ in Definition 3, which is dealt with in Section B.4.4. Theorem B.1 then characterises the asymptotic behaviour of said recursion. In particular, the centred recursion of (S12)-(S15), which is not symmetric, is embedded in a symmetric AMP recursion as defined in B.4.1. A subset of the resulting orbit then corresponds to the orbit of interest, namely (S12)-(S15). This approach was proposed in Javanmard and Montanari (2013, Lemma 1) and also followed in Sur and Candès (2019, Lemma 3) and Donoho and Montanari (2016, Lemma C.1). The symmetric AMP instance is built up for even and odd iterates and the partition $[n] \cup [N] \setminus [n]$ of $[N]$ separately, which is then extended to a complete AMP recursion.

In what follows, iterates are indicated by the letters $t, t-1, t-2$ and 0, which for scalar values are used as subscripts, e.g μ_0, σ_{t-1} and are used as superscripts for functions and nonscalar elements, e.g. $\Psi^{t-1}(\boldsymbol{y}, \boldsymbol{R}^{t-1})$.

Relating a recursion that is parametrised in terms of model parameters $\boldsymbol{y}, y_i, \boldsymbol{X}, \boldsymbol{x}_i$ etc. and a general AMP recursion, which is parametrised by elements such as $x, y, \mathbf{x}_i, \mathbf{y}_i, Y_a$ may lead to confusion. For this reason, an effort is made to point out the disparity. It is specifically pointed out that boldface elements pertaining to the original model use mathematical boldface (e.g. $\boldsymbol{x}_i$), while boldface elements of a AMP recursion use standard boldface (e.g. $\mathbf{x}_i$). Moreover, note that Javanmard and Montanari (2013) chose symbols x, y to denote elements in $\mathcal{V}_{q,N}$, and in particular, $y = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$, a convention which is kept for comparability. This is not the same as $\boldsymbol{y} = (y_1, \ldots, y_n)$, which is used to denote the response variables in the logistic regression model. Moreover, the response variables y_i are assumed to be realizations of random variables Y_i for $i = 1, \ldots, n$. This may stand in conflict with Y_a, the random variables in $\Re^q$ as defined in the AMP setup, which are indexed by by partition cell labels $a \in [q']$. To avoid ill-definedness, the index i is exclusively used when referring to the Bernoulli random variables Y_i that give rise to observations y_i, while the subscript a is used in the context of the random variables for the AMP recursion. Finally, recall that $\boldsymbol{\beta}_0 \in \Re^p$ is the parameter vector that is to be estimated, not an initial condition.

Let q be an arbitrarily large integer, $q' = 2$, $N = n + p$, where n, p denote the rows and columns of $\boldsymbol{X}$, respectively, and $\kappa = \kappa(N) = p/n$. Following Definition 1, 1), let $\boldsymbol{A}$ be the symmetric $N \times N$ matrix whose

upper triangular entries are given by $\boldsymbol{A}_{ij} = \sqrt{n/N}\boldsymbol{X}_{i,j-n}$ for $i \in [n]$ and $j \in [N] \setminus [n]$, and all other entries $\boldsymbol{A}_{ij}$ $i < j$ are i.i.d. $N(0, N^{-1})$ independent of all other entries in $\boldsymbol{A}$. That is, letting $\boldsymbol{B}_1, \boldsymbol{B}_2$ be $n \times n$ matrices with i.i.d. $N(0, (2N)^{-1})$ entries and defining $\boldsymbol{C}_k = \boldsymbol{B}_k + \boldsymbol{B}_k^\top$ for $k = 1, 2$, then

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{C}_1 & \sqrt{\frac{n}{N}}\boldsymbol{X} \\ \sqrt{\frac{n}{N}}\boldsymbol{X}^\top & \boldsymbol{C}_2 \end{bmatrix}. \tag{S19}$$

Next, the collection of functions $\mathcal{F}$ is constructed. For this, recall from Definition 3, 2), that the collections $\mathcal{F}_N$ of a converging sequence of AMP-instances with functions $f^i(\cdot, \cdot)$ may depend on $\mathbf{y}_i$ for some $y \in \mathcal{V}_{q,N}$. Hence, given $\mathbf{x}, \mathbf{y}_i \in \Re^q$, which should be thought of as entries of some $x, y \in \mathcal{V}_{q,N}$ that are concretely specified in (S22) and (S23), for $i \in [n]$, and odd t, define $f^i(\mathbf{x}, t)$ as

$$f^i(\mathbf{x}, t) = \kappa^{-1}\sqrt{\frac{N}{n}}\left[0, \Psi^0\left(h(\mathbf{x}(1), \mathbf{y}_i(1)), \mathbf{x}(3)\right), 0, \Psi^1\left(h(\mathbf{x}(1), \mathbf{y}_i(1)), \mathbf{x}(5)\right), \ldots, \Psi^{\frac{t-1}{2}}\left(h(\mathbf{x}(1), \mathbf{y}_i(1)), \mathbf{x}(t+2)\right), 0, 0, \ldots\right]^\top. \tag{S20}$$

whereas for $i \in [N] \setminus [n]$ and odd t, let $f^i(\mathbf{x}, t) = \mathbf{0}_q$. Further, for $i \in [n]$ and even t, $f^i(\mathbf{x}, t) = \mathbf{0}_q$ and for $i \in [N] \setminus [n]$, even t, let

$$f^i(\mathbf{x}, t) = \sqrt{\frac{N}{n}}\left[\mathbf{y}_i(1), 0, \mathbf{y}_i(2) + \mu_0\mathbf{y}_i(1), 0, \mathbf{x}(2) + \mu_1\mathbf{y}_i(1), 0, \mathbf{x}(4) + \mu_2\mathbf{y}_i(1), 0, \ldots, \mathbf{x}(t) + \mu_{\frac{t}{2}}\mathbf{y}_i(1), 0, 0, \ldots\right]^\top. \tag{S21}$$

For well-definedness of (S20), (S21), which map to $\Re^q$, it must be the case that $t \leq q - 3$, which is of no consequence as q can always be chosen large enough to accommodate any $t \in \Re$. Hence, without loss of generality, all results are stated for $t \in [q]$, rather than $t \in [\tilde{q}]$ for $\tilde{q} = q + 3$ to simplify presentation.

It is straightforward to establish that $f^i(\cdot, t)$ is locally Lipschitz in its first argument, which is shown formally for the smoothed version of B.4.4, to which Theorem B.1 applies ultimately. To conclude the symmetric AMP-instance, let $x^0 = 0_{N,q}$ be the N-vector with elements $\mathbf{0}_q$.

With these definitions in place, it is possible to define the iterates x^t and y that give rise to the target AMP recursion. In particular, let $\boldsymbol{\eta}_0 = \boldsymbol{X}\boldsymbol{\beta}_0$ and for some integer k, $t \in [q]$, define $\mathbf{x}_i^t$ as

$$\mathbf{x}_i^t = \begin{cases} \left[\boldsymbol{\eta}_{0,i}, 0, \boldsymbol{R}_i^0, 0, \boldsymbol{R}_i^1, 0, \ldots, \boldsymbol{R}_i^{\frac{t-1}{2}}, 0, 0, \ldots\right]^\top & \text{if } t = 2k+1, i \in [n] \\ \mathbf{0}_q & \text{if } t = 2k+1, i \in [N] \setminus [n] \\ \mathbf{0}_q & \text{if } t = 2k, i \in [n] \\ \left[0, \boldsymbol{\nu}_{i-n}^1, 0, \boldsymbol{\nu}_{i-n}^2, \ldots, \boldsymbol{\nu}_{i-n}^{\frac{t}{2}}, 0, 0, \ldots\right]^\top & \text{if } t = 2k, i \in [N] \setminus [n] \end{cases} \tag{S22}$$

Finally, let $y \in \mathcal{V}_{N,q}$ have vectors $\mathbf{y}_i \in \Re^q$, $i = 1, \ldots, N$, such that

$$\mathbf{y}_i(k) = \begin{cases} \bar{\varepsilon}_i & \text{if } i \in [n], k = 1 \\ \boldsymbol{\beta}_{0,i-n} & \text{if } i \in [N] \setminus [n], k = 1 \\ \boldsymbol{\nu}_{i-n}^0 & \text{if } i \in [N] \setminus [n], k = 2 \\ 0 & \text{else} \end{cases} \tag{S23}$$

That is, y can be seen as the $q \times N$ matrix

$$y = \begin{bmatrix} \bar{\varepsilon}_1 & \bar{\varepsilon}_2 & \ldots & \bar{\varepsilon}_n & \boldsymbol{\beta}_{0,1} & \boldsymbol{\beta}_{0,2} & \ldots & \boldsymbol{\beta}_{0,p} \\ 0 & 0 & \ldots & 0 & \boldsymbol{\nu}_1^0 & \boldsymbol{\nu}_2^0 & \ldots & \boldsymbol{\nu}_p^0 \\ 0 & & & & \ldots & & & 0 \\ \vdots & & & & \ddots & & & \vdots \\ 0 & & & & \ldots & & & 0 \end{bmatrix}$$

It is now possible to define an AMP-orbit corresponding to the above AMP instance, which is subject of the next two Lemmas. For this, note from Definition 2, equation (S7), that the ith component of an AMP

13

orbit $\{x^t\}_{t\in\mathbb{N}}$ for an AMP instance $(\boldsymbol{A}, \mathcal{F}, x^0)$ is given by

$$\mathbf{x}_i^t = \sum_{j=1}^{N} \boldsymbol{A}_{ij} f^j(\mathbf{x}_j^{t-1}, t-1) - \frac{1}{N} \left\{ \sum_{j=1}^{N} \frac{\partial}{\partial \mathbf{x}} f^j(\mathbf{x}, t-1) \Big|_{\mathbf{x}=\mathbf{x}_j^{t-1}} \right\} f^i(\mathbf{x}_i^{t-2}, t-2). \tag{S24}$$

Lemma B.1. *Consider the symmetric AMP instance $(\boldsymbol{A}, \mathcal{F}, x^0)$ defined by (S19), (S20), (S21) and $x^0 = 0_{N,q}$ and the iterates $\{x^t\}_{t\in[q]}$, the entries of which are defined in (S22) and (S12)-(S15). Then, for $t \in [q]$ being even, and $i \in [N] \setminus [n]$ and $t \in [q]$ odd with $i \in [n]$, these iterates satisfy the recursion of (S24).*

Proof. The proof follows from strong induction on t for even and odd iterates and straightforward algebra.

Base case (i): $t = 1, i \in [n]$. Note that by construction

$$\begin{aligned}
\sum_{j=1}^{N} \boldsymbol{A}_{ij} f^j(\mathbf{x}_j^0, 0) &= \sum_{j=n+1}^{n+p} \boldsymbol{A}_{ij} f^j(\mathbf{x}_j^0, 0) \\
&= \sqrt{\frac{n}{N}} \sum_{j=1}^{p} \boldsymbol{X}_{ij} f^{j+n}(\mathbf{x}_{j+n}^0, 0) \\
&= \sum_{j=1}^{p} \boldsymbol{X}_{ij} \left[\mathbf{y}_{j+n}(1), 0, \mathbf{y}_{j+n}(2) + \mu_0 \mathbf{y}_{j+n}(1), 0, 0, \dots \right]^\top \\
&= \sum_{j=1}^{p} \boldsymbol{X}_{ij} \left[\boldsymbol{\beta}_{0,j}, 0, \boldsymbol{\nu}_j^0 + \mu_0 \boldsymbol{\beta}_{0,j}, 0, 0, \dots \right]^\top \\
&= \sum_{j=1}^{p} \boldsymbol{X}_{ij} \left[\boldsymbol{\beta}_{0,j}, 0, \hat{\boldsymbol{\beta}}_j^0 - \mu_0 \boldsymbol{\beta}_{0,j} + \mu_0 \boldsymbol{\beta}_{0,j}, 0, 0, \dots \right]^\top \\
&= \left[\sum_{j=1}^{p} \boldsymbol{X}_{ij} \boldsymbol{\beta}_{0,j}, 0, \sum_{j=1}^{p} \boldsymbol{X}_{ij} \hat{\boldsymbol{\beta}}_j^0, 0, 0, \dots \right]^\top \\
&= \left[\boldsymbol{\eta}_{0,i}, 0, \boldsymbol{R}_i^0, 0, 0, \dots \right]^\top \\
&= \mathbf{x}_i^1,
\end{aligned} \tag{S25}$$

The first equality follows from the definition of (S21), the second from the definition of (S19), the third again from (S21), the fourth from (S23), the fifth and seventh from the definition of $\boldsymbol{\nu}^0, \boldsymbol{R}^0$ above (S12) and since $\boldsymbol{\eta}_0 = \boldsymbol{X}\boldsymbol{\beta}_0$ and the last from (S22). Concordance with the full recursion of (S24) follows since $f^i(\mathbf{x}_i^{-1}, -1) = \mathbf{0}_q$ by definition.

Base case (ii): $t = 2, i \in [N] \setminus [n]$. Again it is easily checked that

$$\begin{aligned}
\sum_{j=1}^{N} \boldsymbol{A}_{ij} f^j(\mathbf{x}_j^1, 1) &= \sum_{j=1}^{n} \boldsymbol{A}_{ij} f^j(\mathbf{x}_j^1, 1) \\
&= \sqrt{\frac{N}{n}} \sum_{j=1}^{n} \boldsymbol{X}_{ji} f^j(\mathbf{x}_j^1, 1) \\
&= \kappa^{-1} \sum_{j=1}^{n} \boldsymbol{X}_{ji} \left[0, \Psi^0(h(\mathbf{x}_j^1(1), y_j(1)), \mathbf{x}_j^1(3)), 0, 0 \dots \right]^\top \\
&= \kappa^{-1} \sum_{j=1}^{n} \boldsymbol{X}_{ji} \left[0, \Psi^0(h(\boldsymbol{\eta}_{0,j}, \bar{\varepsilon}_j), \boldsymbol{R}_j^0), 0, 0 \dots \right]^\top,
\end{aligned} \tag{S26}$$

14

where the first equality comes from (S20), the second and third from (S19), the fourth again from (S20) and the last from (S25). Now recalling that $y_i^* = h(\boldsymbol{\eta}_{0,i}, \bar{\varepsilon}_i)$ for $h(Z, \bar{\varepsilon}) = \alpha \{1 - \text{sign}(\bar{\varepsilon} - \zeta'(Z))\}/2 + (1-\alpha)/2$, it immediately follows that the nonzero element in the last line of (S26) corresponds to the ith entry of $\kappa^{-1} \boldsymbol{X}^\top \Psi^0(\boldsymbol{y}^*, \boldsymbol{R}^0)$.

Next consider

$$
\sum_{j=1}^N \frac{\partial}{\partial \mathbf{x}} f^j(\mathbf{x}, 1)\Big|_{\mathbf{x}=\mathbf{x}_j^1} = \sum_{j=1}^n \frac{\partial}{\partial \mathbf{x}} f^j(\mathbf{x}, 1)\Big|_{\mathbf{x}=\mathbf{x}_j^1}
$$

$$
= \kappa^{-1} \sqrt{\frac{N}{n}} \sum_{j=1}^n \left[0, \tfrac{\partial}{\partial \mathbf{x}} \Psi^0(h(\mathbf{x}(1), \mathbf{y}(1)), \mathbf{x}(3)), 0, 0, \ldots\right]^\top \Big|_{\mathbf{x}=\mathbf{x}_j^1}
$$

$$
= \kappa^{-1} \sqrt{\frac{N}{n}} \sum_{j=1}^n
\begin{bmatrix}
0 & 0 & \cdots & & \\
\frac{\partial}{\partial Z} \Psi^0(h(Z, \bar{\varepsilon}_j), \boldsymbol{R}_j^0)\big|_{Z=\boldsymbol{\eta}_{0,j}} & 0 & \frac{\partial}{\partial R} \Psi^0(y_j^*, R)\big|_{R=\boldsymbol{R}_j^0} & 0 & 0 & \cdots \\
0 & 0 & \cdots & & \\
\vdots & \ddots & & &
\end{bmatrix},
$$

(S27)

where the first two equalities follow from (S20) and the last from straightforward differentiation and $\mathbf{x}^1$ from (S25).

Finally, it immediately follows from (S21) that

$$
f^i(\mathbf{x}_i^0, 0) = \sqrt{\frac{N}{n}} \left[\mathbf{y}_i(1), 0, \mathbf{y}_i(2) + \mu_0 \mathbf{y}_i(1), 0, 0 \ldots\right]^\top
$$

$$
= \sqrt{\frac{N}{n}} \left[\boldsymbol{\beta}_{0,i-n}, 0, \boldsymbol{\nu}_{i-n}^0 + \mu_0 \boldsymbol{\beta}_{0,i-n}, 0, 0 \ldots\right]^\top
$$

(S28)

so that (S27) and (S28) yield

$$
-\frac{1}{N} \left\{\sum_{j=1}^N \frac{\partial}{\partial \mathbf{x}} f^j(\mathbf{x}, 1)\Big|_{\mathbf{x}=\mathbf{x}_j^1}\right\} f^i(\mathbf{x}_i^0, 0)
$$

$$
= -\frac{1}{n\kappa}
\begin{bmatrix}
0 & 0 & \cdots & & \\
\sum_{j=1}^n \frac{\partial}{\partial Z} \Psi^0(h(Z, \bar{\varepsilon}_j), \boldsymbol{R}_j^0)\big|_{Z=\boldsymbol{\eta}_{0,j}} & 0 & \sum_{j=1}^n \frac{\partial}{\partial R} \Psi^0(y_j^*, R)\big|_{R=\boldsymbol{R}_j^0} & 0 & 0 & \cdots \\
0 & 0 & \cdots & & \\
\vdots & \ddots & & &
\end{bmatrix}
\begin{bmatrix}
\boldsymbol{\beta}_{0,i-n} \\
0 \\
\boldsymbol{\nu}_{i-n}^0 + \mu_0 \boldsymbol{\beta}_{0,i-n} \\
0 \\
0 \\
\vdots
\end{bmatrix}
$$

$$
= \left[0, \left\{-\tfrac{1}{n\kappa} \sum_{j=1}^n \tfrac{\partial}{\partial Z} \Psi^0(h(Z, \bar{\varepsilon}_j), \boldsymbol{R}_j^0)\big|_{Z=\boldsymbol{\eta}_{0,j}}\right\} \boldsymbol{\beta}_{0,i-n} + \left\{-\tfrac{1}{n\kappa} \sum_{j=1}^n \tfrac{\partial}{\partial R} \Psi^0(y_j^*, R)\big|_{R=\boldsymbol{R}_j^0}\right\} \left(\boldsymbol{\nu}_{i-n}^0 + \mu_0 \boldsymbol{\beta}_{0,i-n}\right), 0, 0, \ldots\right]
$$

$$
= \left[0, -a_1 \boldsymbol{\beta}_{0,i-n} + q_0 \left(\boldsymbol{\nu}_{i-n}^0 + \mu_0 \boldsymbol{\beta}_{0,i-n}\right), 0, 0, \ldots\right]^\top, \quad \text{(S29)}
$$

Hence, from (S26), (S29) and (S12), (S14), (S15), it follows that the second entry of (S24) equals

$$
\kappa^{-1} \sum_{j=1}^n \boldsymbol{X}_{ji} \Psi^0(h(\boldsymbol{\eta}_{0,j}, \bar{\varepsilon}_j), \boldsymbol{R}_j^0) - a_1 \boldsymbol{\beta}_{0,i-n} + q_0 \left(\boldsymbol{\nu}_{i-n}^0 + \mu_0 \boldsymbol{\beta}_{0,i-n}\right) = \boldsymbol{\nu}_{i-n}^1,
$$

while all other entries are zero. Hence, by construction of (S22) the second base case holds.

Induction hypothesis: Assume that there exists some $t < q$, $t \geq 2$ such that for all $t' \leq t$ if t' is even, and $i \in [N] \setminus [n]$ or if t' is odd with $i \in [n]$, the iterates of (S22) satisfy the recursion of (S24).

15

Induction step (i): $t + 1$ odd, $i \in [n]$. Most derivations are analogous to Base case (i), so that details are skipped on repetitive steps. First note that due to the nested nature of the iterates $\mathbf{x}_i^t$ in (S22) and the induction hypothesis, the induction step must only be established for the last nonzero entry in $\mathbf{x}_j^{t+1}$.

For this, note that

$$\sum_{j=1}^{N} \boldsymbol{A}_{ij} f^j(\mathbf{x}_j^t, t) = \sum_{j=1}^{p} \boldsymbol{X}_{ij} \left[\boldsymbol{\beta}_{0,j}, 0, \boldsymbol{\nu}_j^0 + \mu_0 \boldsymbol{\beta}_{0,j}, 0, \boldsymbol{\nu}_j^1 + \mu_1 \boldsymbol{\beta}_{0,j}, 0, \ldots, \boldsymbol{\nu}_j^{\frac{t}{2}} + \mu_{\frac{t}{2}} \boldsymbol{\beta}_{0,j}, 0, 0, \ldots \right]^\top . \tag{S30}$$

Next, it is easily seen that

$$\sqrt{\frac{n}{N}} \sum_{j=1}^{N} \frac{\partial}{\partial \mathbf{x}} f^j(\mathbf{x}, t) \Big|_{\mathbf{x}=\mathbf{x}_j^t} = \sqrt{\frac{n}{N}} \sum_{j=1}^{p} \frac{\partial}{\partial \mathbf{x}} f^{n+j}(\mathbf{x}, t) \Big|_{\mathbf{x}=\mathbf{x}_{n+j}^t}$$

$$= \sum_{j=1}^{p} \frac{\partial}{\partial \mathbf{x}} \left[\mathbf{y}_i(1), 0, \mathbf{y}_i(2) + \mu_0 \mathbf{y}_i(1), 0, \mathbf{x}(2) + \mu_1 \mathbf{y}_i(1), 0, \mathbf{x}(4) + \mu_2 \mathbf{y}_i(1), 0, \ldots, \mathbf{x}(t) + \mu_{\frac{t}{2}} \mathbf{y}_i(1), 0, \ldots \right]^\top \Big|_{\mathbf{x}=\mathbf{x}_{n+j}^t}$$

$$= \sum_{j=1}^{p} \tilde{I}_{q \times q} = p \tilde{I}_{q \times q}, \tag{S31}$$

where $\tilde{I}_{q \times q}$ whose odd rows with row index $5 \le i \le t + 3$ have the $(i-3)$rd element equal to one, while all other elements of this matrix are equal to zero.

Further, note that for $t + 1$ odd, $i \in [n]$,

$$\kappa \sqrt{\frac{n}{N}} f^i(\mathbf{x}_i^{t-1}, t-1) = \begin{bmatrix} 0 \\ \Psi^0 \left(h(\mathbf{x}_i^{t-1}(1), \mathbf{y}_i(1)), \mathbf{x}_i^{t-1}(3) \right) \\ 0 \\ \Psi^1 \left(h(\mathbf{x}_i^{t-1}(1), \mathbf{y}_i(1)), \mathbf{x}_i^{t-1}(5) \right) \\ 0 \\ \vdots \\ \Psi^{\frac{t-2}{2}} \left(h(\mathbf{x}_i^{t-1}(1), \mathbf{y}_i(1)), \mathbf{x}_i^{t-1}(t+1) \right) \\ 0 \\ \vdots \end{bmatrix} = \begin{bmatrix} 0 \\ \Psi^0 \left(h(\boldsymbol{\eta}_{0,i}, \bar{\varepsilon}_i), \boldsymbol{R}_i^0 \right) \\ 0 \\ \Psi^1 \left(h(\boldsymbol{\eta}_{0,i}, \bar{\varepsilon}_i), \boldsymbol{R}_i^1 \right) \\ 0 \\ \vdots \\ \Psi^{\frac{t}{2}-1} \left(h(\boldsymbol{\eta}_{0,i}, \bar{\varepsilon}_i), \boldsymbol{R}_i^{\frac{t}{2}-1} \right) \\ 0 \\ \vdots \end{bmatrix} , \tag{S32}$$

by definition of (S21), (S22) and (S23). Hence, from (S31) and (S32), it follows that

$$-\frac{1}{N} \left\{ \sum_{j=1}^{N} \frac{\partial}{\partial \mathbf{x}} f^j(\mathbf{x}, t) \Big|_{\mathbf{x}=\mathbf{x}_j^t} \right\} f^i(\mathbf{x}_i^{t-1}, t-1) = -\kappa^{-1} \frac{p}{n} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \Psi^0 \left(h(\boldsymbol{\eta}_{0,i}, \bar{\varepsilon}_i), \boldsymbol{R}_i^0 \right) \\ 0 \\ \Psi^1 \left(h(\boldsymbol{\eta}_{0,i}, \bar{\varepsilon}_i), \boldsymbol{R}_i^1 \right) \\ 0 \\ \vdots \\ \Psi^{\frac{t}{2}-1} \left(h(\boldsymbol{\eta}_{0,i}, \bar{\varepsilon}_i), \boldsymbol{R}_i^{\frac{t}{2}-1} \right) \\ 0 \\ \vdots \end{bmatrix} ,$$

Hence, it follows from equations (S30) and (S30), that the last nonzero entry of

$$\sum_{j=1}^{N} \boldsymbol{A}_{ij} f^j(\mathbf{x}_j^t, t) - \frac{1}{N} \left\{ \sum_{j=1}^{N} \frac{\partial}{\partial \mathbf{x}} f^j(\mathbf{x}, t) \Big|_{\mathbf{x}=\mathbf{x}_j^t} \right\} f^i(\mathbf{x}_i^{t-1}, t-1)$$

is given by

$$\sum_{j=1}^{p} \boldsymbol{X}_{ij} \left(\boldsymbol{\nu}_j^{\frac{t}{2}} + \mu_{\frac{t}{2}} \boldsymbol{\beta}_{0,j} \right) - \Psi^{\frac{t}{2}-1} \left(h(\boldsymbol{\eta}_{0,i}, \bar{\varepsilon}_i), \boldsymbol{R}^{\frac{t}{2}-1} \right),$$

which by (S13) corresponds to the ith entry of $\boldsymbol{R}^{\frac{t}{2}}$ and that corresponds precisely to the last nonzero entry of $\mathbf{x}_i^{t+1}$ as defined in (S22). This establishes the first part of the induction step.

Induction step (ii): $t+1$ even, $i \in [N] \setminus [n]$. This step is completely analogous to base case (ii) and therefore omitted.

By strong induction, the claim of the Lemma holds for any $t \in [q]$. $\qquad\square$

With this result it is possible to fill up the missing parts the incomplete recursion of Lemma B.1. Hence, define iterate $\tilde{x}^t$ as follows. Let $\tilde{x}^0 = 0_{N,q}$ and for $t \in [q]$ odd and $i \in [n]$ or if t is even and $i \in [N] \setminus [n]$, set $\tilde{\mathbf{x}}_i^t = \mathbf{x}_i^t$. Otherwise, let

$$\tilde{\mathbf{x}}_i^t = \sum_{j=1}^{N} \boldsymbol{A}_{ij} f^j(\tilde{\mathbf{x}}_j^{t-1}, t-1) - \frac{1}{N} \left\{ \sum_{j=1}^{N} \frac{\partial}{\partial \mathbf{x}} f^j(\tilde{\mathbf{x}}, t-1) \Big|_{\tilde{\mathbf{x}}=\tilde{\mathbf{x}}_j^{t-1}} \right\} f^i(\tilde{\mathbf{x}}_i^{t-2}, t-2), \qquad (\text{S33})$$

again with the convention that terms involving negative t-indices are zero. The Lemma below establishes that this recursion is indeed a AMP orbit corresponding to instance $(\boldsymbol{A}, \mathcal{F}, \tilde{x}^0)$ and that the orbit of interest, i.e. the iterates of Lemma B.1, are preserved.

Lemma B.2. *Given the AMP instance $(\boldsymbol{A}, \mathcal{F}, \tilde{x}^0)$ of Lemma B.1, the recursion of (S33) is an orbit according to Definition 2. Further, for $t \in [q]$ odd and $i \in [n]$ or if t is even and $i \in [N] \setminus [n]$, the iterates $\tilde{\mathbf{x}}_i^t$ correspond to those in Lemma B.1.*

Proof. The statement of the Lemma is shown by strong induction on t.

Base case (i): $t = 1$. Since $\tilde{x}^0 = x^0 = 0_{n,q}$, by base case (i) of Lemma B.1, the second part of the Lemma holds for $t = 1$. For $i \in [n]$, the first part of the claim follows from base case (i) of Lemma B.1 too, while for $i \in [N] \setminus [n]$ it follows trivially from (S33).

Base case (ii) $t = 2$. Again, note that by the base case (ii) in Lemma B.1, it the claim of the Lemma immediately follows for $i \in [N] \setminus [n]$. For $i \in [n]$, the second part of the Lemma follows from (S33).

Induction hypothesis: The claim of the Lemma holds for all t' up to some $t < q$, $t \geq 2$.

Induction step (i): $t+1$ odd. Consider $i \in [n]$. From the induction hypothesis, one hast that $\tilde{\mathbf{x}}_j^t = \mathbf{x}_j^t$ for $i \in [N] \setminus [n]$ and similarly $\tilde{\mathbf{x}}_j^{t-1} = \mathbf{x}_j^{t-1}$ for $i \in [n]$. Hence again, the corresponding induction step of Lemma B.1 shows that the recursion an orbit according to Definition 2 and that the iterates $\tilde{\mathbf{x}}_i^t$ correspond to those in Lemma B.1. Now consider $i \in [N] \setminus [n]$. The induction step holds trivially by construction of (S33).

Induction step (ii): $t+1$ even. This case is handled similarly to induction step (i) and is therefore omitted.

By strong induction, the Lemma is proven for all $t \in [q]$. $\qquad\square$

Hence, if $(\boldsymbol{A}, \mathcal{F}, \tilde{x}^0)$ were shown to be a converging AMP instance according to Definition 3, then Theorem B.1 along with the state evolution of Definition 4, would give the large sample behaviour of the corresponding orbit of interest. However, as remarked in Feng *et al.* (2022), Section 4.7, the function g in Definition 3 cannot be Lipschitz, which is owed to the indicator function $h(Z, \bar{\varepsilon})$, so that Theorem B.1 does not apply. The next section deals with this technicality.

B.4.4 A smooth approximation

This section introduces a smooth approximation to the original recursions (S10)-(S11) and to (S12)-(S15), the latter of which can be used in a matrix-valued AMP embedding for which Theorem B.1 applies. A limiting argument then shows that the theorem still applies in the limit where the smooth approximation approaches the nonsmooth original recursion.

In particular, the step function $\{1 - \text{sign}(x)\}/2$ can, amongst others, be approximated by $\left\{1 - \frac{x}{|x|+\epsilon}\right\}/2$ for some positive constant ϵ. Note that defining the Bernoulli response $Y = \{1 - \text{sign}(\bar{\varepsilon} - \zeta'(Z))\}/2$ rather than the more conventional $Y = \mathbb{1}\{\bar{\varepsilon} < \zeta'(Z)\}$ comes without loss of generality, as per the modelling assumption, $\boldsymbol{\eta}_{0,i}$ are independent $N(0,\gamma^2)$ normal random variables and $\bar{\varepsilon}_i$ are independent of $\boldsymbol{\eta}_0$ and uniformly distributed on the unit interval so that the event $\{\exists i \in [n] : \bar{\varepsilon}_i = \zeta'(\boldsymbol{\eta}_{0,i})\}$ has measure zero.

Hence, in what follows, denote by $h_\epsilon(Z, \bar{\varepsilon})$ the function

$$h_\epsilon(Z, \bar{\varepsilon}) = \frac{\alpha}{2}\left\{1 - \frac{\bar{\varepsilon} - \zeta'(Z)}{|\bar{\varepsilon} - \zeta'(Z)| + \epsilon}\right\} + \frac{1-\alpha}{2}\,.$$

Consequently, define new sets of recursions analogous to (S10)-(S11) and (S12)-(S15) where all appearances of $h(\cdot, \cdot)$ are replaced by $h_\epsilon(\cdot, \cdot)$. In particular, define the approximate target AMP recursion with initial conditions $\hat{\boldsymbol{\beta}}^0$, $\boldsymbol{S}^0 = \boldsymbol{X}\hat{\boldsymbol{\beta}}^0$, and a sequence of constants $\{_\epsilon\mu_t, {}_\epsilon b_t, {}_\epsilon\sigma_t\}_{t\in\mathbb{N}_0}$, for $t \in \mathbb{N}$, as

$$_\epsilon\hat{\boldsymbol{\beta}}^t = {}_\epsilon\hat{\boldsymbol{\beta}}^{t-1} + \kappa^{-1}\boldsymbol{X}^\top\Psi^{t-1}(_\epsilon\boldsymbol{y}^*, {}_\epsilon\boldsymbol{S}^{t-1}) \tag{S34}$$

$$_\epsilon\boldsymbol{S}^t = \boldsymbol{X}_\epsilon\hat{\boldsymbol{\beta}}^t - \Psi^{t-1}(_\epsilon\boldsymbol{y}^*, {}_\epsilon\boldsymbol{S}^{t-1})\,, \tag{S35}$$

where $\Psi^t(_\epsilon\boldsymbol{y}^*, {}_\epsilon\boldsymbol{S}^t)$ acts elementwise on $_\epsilon\boldsymbol{y}^*, {}_\epsilon\boldsymbol{S}^t$ as

$$\Psi^t(_\epsilon y_i^*, {}_\epsilon\boldsymbol{S}_i^t) = {}_\epsilon b_t\left\{_\epsilon y_i^* - \zeta'\left(\text{prox}_{_\epsilon b_t\zeta}\left(_\epsilon b_t\epsilon y_i^* + {}_\epsilon\boldsymbol{S}_i^t\right)\right)\right\}\,,$$

for $i = 1, \ldots, n$ and where $_\epsilon y_i^* = \alpha h_\epsilon(\boldsymbol{\eta}_{0,i}, \bar{\varepsilon}_i) + (1-\alpha)/2$. Further, the state evolution is determined by the system of nonlinear equations

$$1 - \kappa = \text{E}\left[\frac{1}{1 + {}_\epsilon b_t\zeta''\left(\text{prox}_{_\epsilon b_t\zeta}\left(_\epsilon b_{t\epsilon}Y^* + {}_\epsilon Z_t\right)\right)}\right] \tag{S36}$$

$$_\epsilon\mu_{t+1} = \frac{_\epsilon b_t}{\kappa\gamma^2}\text{E}\left[Z\left\{_\epsilon Y^* - \zeta'\left(\text{prox}_{_\epsilon b_t\zeta}\left(_\epsilon b_{t\epsilon}Y^* + {}_\epsilon Z_t\right)\right)\right\}\right] + {}_\epsilon\mu_t \tag{S37}$$

$$_\epsilon\sigma_{t+1}^2 = \frac{_\epsilon b_t^2}{\kappa^2}\text{E}\left[\left\{_\epsilon Y^* - \zeta'\left(\text{prox}_{_\epsilon b_t\zeta}\left(_\epsilon b_{t\epsilon}Y^* + {}_\epsilon Z_t\right)\right)\right\}^2\right]\,, \tag{S38}$$

for independent $Z, \tilde{Z} \sim N(0, \gamma^2)$, $G \sim N(0,1)$, and $\bar{\varepsilon} \sim U(0,1)$, and letting

$$_\epsilon Y^* = h_\epsilon(Z, \bar{\varepsilon}), \quad {}_\epsilon Z_t = {}_\epsilon\mu_t Z + \kappa^{1/2}{}_\epsilon\sigma_t G\,.$$

Given initial conditions $\boldsymbol{\nu}^0 = \hat{\boldsymbol{\beta}}^0 - \mu_0\boldsymbol{\beta}_0$, $\boldsymbol{R}^0 = \boldsymbol{S}^0$ for $\hat{\boldsymbol{\beta}}^0, \boldsymbol{\beta}_0 \in \Re^p$, $\mu_0 \in \Re$ and $\boldsymbol{S}^0 \in \Re^n$, for $t \in \mathbb{N}$, the approximate centred recursion of (S12)-(S15) is defined as

$$_\epsilon\boldsymbol{\nu}^t = {}_\epsilon q_{t-1}(_\epsilon\boldsymbol{\nu}^{t-1} + {}_\epsilon\mu_{t-1}\boldsymbol{\beta}_0) - {}_\epsilon a_t\boldsymbol{\beta}_0 + \kappa^{-1}\boldsymbol{X}^\top\Psi^{t-1}(_\epsilon\boldsymbol{y}^*, {}_\epsilon\boldsymbol{R}^{t-1}) \tag{S39}$$

$$_\epsilon\boldsymbol{R}^t = \boldsymbol{X}(_\epsilon\boldsymbol{\nu}^t + {}_\epsilon\mu_t\boldsymbol{\beta}_0) - \Psi^{t-1}(_\epsilon\boldsymbol{y}^*, {}_\epsilon\boldsymbol{R}^{t-1})\,, \tag{S40}$$

where $a_0 = \mu_0$ and for $t \in \mathbb{N}$,

$$_\epsilon q_t = -\frac{1}{\kappa n}\sum_{i=1}^n\frac{\partial\Psi^t}{\partial R}(_\epsilon y_i^*, R)\Big|_{R={}_\epsilon\boldsymbol{R}_i^t}$$

$$_\epsilon a_t = \frac{1}{\kappa n}\sum_{i=1}^n\frac{\partial\Psi^{t-1}}{\partial Z}(h_\epsilon(Z, \bar{\varepsilon}_i), {}_\epsilon\boldsymbol{R}_i^{t-1})\Big|_{Z=\boldsymbol{x}_i^\top\boldsymbol{\beta}_0}\,, \tag{S41}$$

where $\bar{\varepsilon}_i \sim U(0,1)\,(i=1,\ldots,n)$ are i.i.d. draws independent of all other random variables.

Similarly, define $f^i_{\epsilon}(\mathbf{x},t)$ as functions (S20) and (S21) with $h(Z,\bar{\varepsilon})$ replaced by $h_{\epsilon}(Z,\bar{\varepsilon})$ and denote by $_{\epsilon}\mathbf{x}^t_i$ the orbit of (S22) whose definition is unchanged but whose entries now depend on ϵ through the replacement of $h(\cdot,\cdot)$ by $h_{\epsilon}(\cdot,\cdot)$ in (S12)-(S15). Finally, denote by $_{\epsilon}\tilde{\mathbf{x}}^t_i$ the corresponding full AMP recursion as defined in (S33) with $\mathbf{x}^t_i$ replaced by $_{\epsilon}\mathbf{x}^t_i$. Completely analogous to Lemmas B.1 and B.2, it can be shown that $_{\epsilon}\mathbf{x}^t_i$ is a AMP orbit to the AMP instance $(\boldsymbol{A},\mathcal{F}_{\epsilon},\tilde{x}^0)$, where the definition of $\boldsymbol{A}$ remains unchanged, $\mathcal{F}_{\epsilon}$ is the collection of functions f^i_{ϵ} and $\tilde{x}^0 = \mathbf{0}_{q,N}$ as before. This is easily verified by noting that the proofs of these Lemmas do not depend on the form of the indicator function but are concerned with the general structure of the recursion.

It remains to construct a converging AMP instance following Definition 3 to motivate Theorem B.1. Following the structure of Definition 3,

i) Let q be some arbitrarily large integer, and $q' = 2$

ii) Define $g_{\epsilon} : \Re^q \times \Re^q \times [q'] \times \mathbb{N} \to \Re^q$ as follows. Given $k \in \mathbb{N}_0$,

$$
g_{\epsilon}(\mathbf{x},\mathbf{y},1,2k+1)
$$
$$
= \kappa^{-1}\sqrt{\frac{N}{n}}\left[0,\Psi^0\left(h_{\epsilon}(\mathbf{x}(1),\mathbf{y}(1)),\mathbf{x}(3)\right),0,\Psi^1\left(h_{\epsilon}(\mathbf{x}(1),\mathbf{y}(1)),\mathbf{x}(5)\right),\ldots,\Psi^{\frac{t-1}{2}}\left(h_{\epsilon}(\mathbf{x}(1),\mathbf{y}(1)),\mathbf{x}(t+2)\right),0,0,\ldots\right]^{\top},
$$
$$(S42)$$

and

$$
g_{\epsilon}(\mathbf{x},\mathbf{y},2,2k)
$$
$$
= \sqrt{\frac{N}{n}}\left[\mathbf{y}(1),0,\mathbf{y}(2)+\mu_0\mathbf{y}(1),0,\mathbf{x}(2)+\mu_1\mathbf{y}(1),0,\mathbf{x}(4)+\mu_2\mathbf{y}(1),0,\ldots,\mathbf{x}(t)+\mu_{\frac{t}{2}}\mathbf{y}(1),0,0,\ldots\right]^{\top},
$$
$$(S43)$$

while for all other configurations of t,a, $g_{\epsilon}(\mathbf{x},\mathbf{y},a,t) = \mathbf{0}_q$.

iii) Let $\pi_{\bar{\beta}}$ be the probability distribution such that the empirical distribution of $\boldsymbol{\beta}_0$ converges weakly to $\pi_{\bar{\beta}}$, which is assumed to exist. Further, given initial conditions $\hat{\boldsymbol{\beta}}^0, \mu_0$, let $\pi_{\bar{\nu}}$ be the probability distribution such that the empirical distribution of $\boldsymbol{\nu}^0 = \hat{\boldsymbol{\beta}}^0 - \mu_0\boldsymbol{\beta}_0$ converges weakly to, which is also assumed to exist, such as the joint distribution $\pi_{\bar{\beta},\bar{\nu}}$ which is the limit of the joint distribution of $\boldsymbol{\beta}_0, \boldsymbol{\nu}^0$. Finally, let $\pi_{\bar{\varepsilon}}$ be the distribution of a $U(0,1)$ random variable. Then let P_1 be the distribution of q independent random variables whose marginals are $\pi_{\bar{\varepsilon}}$ for the first coordinate and point masses at zero for all others. Further, let P_2 be the distribution of q random variables whose first two coordinates are distributed according to $\pi_{\bar{\beta},\bar{\nu}}$ and are independent of all other coordinates, whose distributions are independent of another and point masses at zero.

iv) For $N = n+p$, partition $[N]$ into $C^N_1 = [n]$ and $C^N_2 = [N] \setminus [n]$

v) For $a = 1,2$, let
$$
\widehat{\boldsymbol{\Sigma}}^0_a = \plim_{N\to\infty}\frac{1}{|C^N_a|}\sum_{i\in C^N_a} g_{\epsilon}(\mathbf{0}_q,\mathbf{y}_i,a,0)g_{\epsilon}(\mathbf{0}_q,\mathbf{y}_i,a,0)^{\top},
$$

which is assumed to exist. Here, $c_1 = \lim_{N\to\infty} n/(n+p) = 1/(1+\kappa)$ and $c_2 = \lim_{N\to\infty} p/(n+p) = \kappa/(1+\kappa)$ in concordance with Definition 3 1).

Condition iii) appears a bit abstract in its stated generality. For initial condition of relevance in this paper, namely the oracle initialiser $\mu_*\boldsymbol{\beta}_0 + \sigma_*\boldsymbol{\xi}$, where $\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{I}_p)$ independent of everything else, it is readily seen that the condition holds whenever the empirical distribution of $\boldsymbol{\beta}_0$ converges to some distribution $\pi_{\bar{\beta}}$.

To conclude the definition, it must now be shown that for all $r \in [q]$, the rth component of $g_{\epsilon}(\cdot,\cdot,a,t)$ is Lipschitz in its first two arguments and that points 1)-3) of Definition 3 hold. This is subject of the Lemma below.

19

Lemma B.3. *The AMP instance defined above is a converging AMP instance according to Definition 3.*

Proof. It must be shown that 0) each component of $g_\epsilon(\cdot, \cdot, a, t)$ is Lipschitz in its first two arguments and that points 1)-3) of Definition 3 hold. Each point is shown separately.

0) Lipschitzianity of $g_\epsilon(\cdot, \cdot, a, t)$: It is shown in Lemma G.2 ii) that $\Psi^{\frac{t-1}{2}}(h_\epsilon(\mathbf{x}(1), \mathbf{y}(1)), \mathbf{x}(t+2))$ is Lipschitz in $[\mathbf{x}(1), \mathbf{x}(t+2), \mathbf{y}(1)]^\top$. Hence by (S42), for any $k \in \mathbb{N}_0$, $a = 1$, $r \in [q]$, the rth component of $g_\epsilon(\cdot, \cdot, 1, 2k+1)$ is Lipschitz. On the other hand, by (S43), Lipschitzianity of $g_\epsilon(\cdot, \cdot, 2, 2k)$ is immediate since it is an affine transformation of its arguments. For all other combinations of t, a, $g_\epsilon(\cdot, \cdot, a, t) = \mathbf{0}_q$ and thus trivially Lipschitz.

1) Convergence of $|C_a^N|/N$: By definition of $N = n + p$, and $C_N^1 = [n]$, $C_N^2 = [N] \setminus [n]$. Hence, since by assumption, $\kappa(N) = p/n$ converges to some $\kappa \in (0, 1)$, 1) holds.

2): Agreement of f_ϵ^i with the ith component of g_ϵ holds by construction of y in (S23), f_ϵ^i in (S20),(S21), g_ϵ in (S42) and (S43) along with the partition C_N^1, C_N^2. Convergence of the empirical distribution of $\{\mathbf{y}_i\}_{i \in C_N^a}$ to P_a follows by definition of P_a.

3) Initial condition of state evolution: Follows by assumption v). □

It thus follows from Lemma B.2 applied to the smooth approximation ${}_\epsilon\tilde{\mathbf{x}}_i^t$ and Lemma B.3 that $(\boldsymbol{A}, \mathcal{F}_\epsilon, \tilde{x}^0)$ along with iterates ${}_\epsilon\tilde{\mathbf{x}}_i^t$ is a converging AMP instance and Theorem B.1 applies.

Theorem B.2. *Under assumptions (i)-(v), for any pseudo Lipschitz-function of order 2, and for all $t \in [q]$, $a = 1, 2$, almost surely*

$$\lim_{N \to \infty} \frac{1}{|C_a^N|} \sum_{i \in C_a^N} \psi({}_\epsilon\tilde{\mathbf{x}}_i^t, \mathbf{y}) = E\left[\psi({}_\epsilon Z_a^t, Y_a)\right] ,$$

where ${}_\epsilon Z_a^t \sim N(0, {}_\epsilon\boldsymbol{\Sigma}^t)$, and ${}_\epsilon\boldsymbol{\Sigma}^t$ denotes the covariance matrices from state evolution (S8) and (S9) with function $g_\epsilon(\mathbf{x}, \mathbf{y}, a, t)$

Proof. By Lemma B.3 and Theorem B.1, for any pseudo Lipschitz function ψ of order 2,

$$\lim_{N \to \infty} \frac{1}{|C_a^N|} \sum_{i \in C_a^N} \psi({}_\epsilon\tilde{\mathbf{x}}_i^t, \mathbf{y}) = E\left[\psi({}_\epsilon Z_a^t, Y_a)\right] ,$$

where ${}_\epsilon Z_a^t \sim N(0, {}_\epsilon\boldsymbol{\Sigma}^t)$ as defined in the statement of the Theorem. □

Corollary B.2.1. *Under assumptions (i)-(v), and initial conditions $\hat{\boldsymbol{\beta}}^0$,*

$$_\epsilon\mu_0 = \frac{1}{\gamma^2} \lim_{n \to \infty} \frac{\langle \hat{\boldsymbol{\beta}}^0, \boldsymbol{\beta}_0 \rangle}{n}$$

$$_\epsilon\sigma_0^2 = \lim_{n \to \infty} \frac{\left\|\hat{\boldsymbol{\beta}}^0 - \mu_0\boldsymbol{\beta}_0\right\|_2^2}{p} ,$$

for any pseudo Lipschitz-function of order 2, and for all $t \in [q]$, almost surely

$$\lim_{N \to \infty} \frac{1}{p} \sum_{j=1}^{p} \psi\left({}_\epsilon\boldsymbol{\nu}_j^t, \boldsymbol{\beta}_{0,j}\right) = E\left[\psi({}_\epsilon\sigma_t^2 G, \bar{\beta})\right] , \quad \lim_{N \to \infty} \frac{1}{n} \sum_{j=1}^{n} \psi\left(\begin{bmatrix} \boldsymbol{\eta}_{0,j} \\ {}_\epsilon\boldsymbol{R}_i^t \end{bmatrix}, \begin{bmatrix} \bar{\varepsilon}_j \\ 0 \end{bmatrix}\right) = E\left[\psi\left(\begin{bmatrix} Z_1^t \\ {}_\epsilon Z_2^t \end{bmatrix}, \begin{bmatrix} \bar{\varepsilon} \\ 0 \end{bmatrix}\right)\right] ,$$

Proof. The proof consists of two parts, which shall be shown in turn: i) For fixed $\epsilon > 0$, extract the iterates $_\varepsilon\boldsymbol{\nu}^t, {}_\varepsilon\boldsymbol{R}^t$ of (S39) and (S40) from $_\epsilon\tilde{\mathbf{x}}_i^t$ and obtain their asymptotic variances according to Theorem B.2, and ii) relate these variances to the state evolution of (S36)-(S38).

The first step is similar to the derivations in the proof of Lemma 6 in the supplementary material of Sur and Candès (2019). For $t = 1$, it is easily seen from the AMP instance defined in B.4.4 and the state evolutions (S8) and (S9) that

$$
\begin{bmatrix} _\epsilon\boldsymbol{\Sigma}_{1,1}^1 & _\epsilon\boldsymbol{\Sigma}_{1,3}^1 \\ _\epsilon\boldsymbol{\Sigma}_{3,1}^1 & _\epsilon\boldsymbol{\Sigma}_{3,3}^1 \end{bmatrix} = \begin{bmatrix} \lim\limits_{N\to\infty} \frac{\|\boldsymbol{\beta}_0\|_2^2}{n} & \lim\limits_{N\to\infty} \frac{\langle \hat{\boldsymbol{\beta}}^0, \boldsymbol{\beta}_0 \rangle}{n} \\ \lim\limits_{N\to\infty} \frac{\langle \hat{\boldsymbol{\beta}}^0, \boldsymbol{\beta}_0 \rangle}{n} & \lim\limits_{N\to\infty} \frac{\|\hat{\boldsymbol{\beta}}^0\|_2^2}{n} \end{bmatrix} , \tag{S44}
$$

where $_\epsilon\boldsymbol{\Sigma}_{i,j}^1$ denotes the jth entry in the ith row of $_\epsilon\boldsymbol{\Sigma}^1$. All other entries of $_\epsilon\boldsymbol{\Sigma}^1$ are zero. Now from (S22) it follows that for $i \in [n]$, the first and third entries of $_\epsilon\tilde{\mathbf{x}}_i^1$ are $\boldsymbol{\eta}_{0,i}$, $_\epsilon\boldsymbol{R}_i^0 = \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}^0$. Hence, by Theorem B.2,

$$
\lim_{N\to\infty} \frac{1}{n} \sum_{j=1}^n \psi\left(\begin{bmatrix} \boldsymbol{\eta}_{0,j} \\ _\epsilon\boldsymbol{R}_i^0 \end{bmatrix}, \begin{bmatrix} \bar{\varepsilon}_j \\ 0 \end{bmatrix} \right) = \mathrm{E}\left[\psi\left(\boldsymbol{Z}^1, \begin{bmatrix} \bar{\varepsilon} \\ 0 \end{bmatrix} \right) \right],
$$

where $\boldsymbol{Z}^1 = [\boldsymbol{Z}_1^1, \boldsymbol{Z}_2^1]^\top \perp \bar{\varepsilon}$ is bivariate normal with mean zero and covariance (S44).

Next, the only nonzero entry of $_\epsilon\boldsymbol{\Sigma}^2$ is readily computed as $\kappa^{-2}\mathrm{E}\left[\left\{ \Psi^0(h_\epsilon(\boldsymbol{Z}_1^1, \bar{\varepsilon}), \boldsymbol{Z}_2^1) \right\}^2 \right]$. By construction of $_\epsilon\tilde{\mathbf{x}}_i^t$ it follows from (S22) and Lemma B.3 and Theorem B.1 that almost surely

$$
\lim_{N\to\infty} \frac{1}{p} \sum_{j=1}^p \psi(_\epsilon\boldsymbol{\nu}_j^1, \boldsymbol{\beta}_{0,j}) = \mathrm{E}\left[\psi(_\epsilon\tau_1^2 G, \bar{\beta}) \right],
$$

where $_\epsilon\tau_1^2 = \kappa^{-2}\mathrm{E}\left[\left\{ \Psi^0(h_\epsilon(\boldsymbol{Z}_1^1, \bar{\varepsilon}), \boldsymbol{Z}_2^1) \right\}^2 \right]$ and $G \sim N(0,1) \perp \bar{\beta}$. Next consider $_\epsilon\boldsymbol{\Sigma}^3$. By the nested nature of the iterates $_\epsilon\mathbf{x}_i^t$, it follows that the first 3×3 principal submatrix is exactly the one given in (S44). Further, it is readily verified that the remaining nonzero entries are given by

$$
_\epsilon\boldsymbol{\Sigma}_{1,5}^3 = {}_\epsilon\mu_1\gamma^2, \quad _\epsilon\boldsymbol{\Sigma}_{3,5}^3 = {}_\epsilon\mu_1 \lim_{N\to\infty} \frac{\langle \hat{\boldsymbol{\beta}}^0, \boldsymbol{\beta}_0 \rangle}{n}, \quad _\epsilon\boldsymbol{\Sigma}_{5,5}^3 = {}_\epsilon\tau_1^2\kappa + {}_\epsilon\mu_1^2\gamma^2,
$$

where $_\epsilon\mu_1$ comes from recursion (S37) with initial condition μ_0. Again, Lemma B.2 and Theorem B.1 yield almost surely

$$
\lim_{N\to\infty} \frac{1}{n} \sum_{j=1}^n \psi\left(\begin{bmatrix} \boldsymbol{\eta}_{0,j} \\ _\epsilon\boldsymbol{R}_j^0 \\ _\epsilon\boldsymbol{R}_j^1 \end{bmatrix}, \begin{bmatrix} \bar{\varepsilon}_j \\ 0 \\ 0 \end{bmatrix} \right) = \mathrm{E}\left[\psi\left(\boldsymbol{Z}^3, \begin{bmatrix} \bar{\varepsilon} \\ 0 \\ 0 \end{bmatrix} \right) \right],
$$

where $\boldsymbol{Z}^3 = [\boldsymbol{Z}_1^3, \boldsymbol{Z}_2^3, \boldsymbol{Z}_3^3]^\top \perp \bar{\varepsilon}$ is trivariate normal with mean zero and covariance matrix

$$
\begin{bmatrix} _\epsilon\boldsymbol{\Sigma}_{1,1}^3 & _\epsilon\boldsymbol{\Sigma}_{1,3}^3 & _\epsilon\boldsymbol{\Sigma}_{1,5}^3 \\ _\epsilon\boldsymbol{\Sigma}_{3,1}^3 & _\epsilon\boldsymbol{\Sigma}_{3,3}^3 & _\epsilon\boldsymbol{\Sigma}_{3,5}^3 \\ _\epsilon\boldsymbol{\Sigma}_{5,1}^3 & _\epsilon\boldsymbol{\Sigma}_{5,3}^3 & _\epsilon\boldsymbol{\Sigma}_{5,5}^3 \end{bmatrix} = \begin{bmatrix} \lim\limits_{N\to\infty} \frac{\|\boldsymbol{\beta}_0\|_2^2}{n} & \lim\limits_{N\to\infty} \frac{\langle \hat{\boldsymbol{\beta}}^0, \boldsymbol{\beta}_0 \rangle}{n} & _\epsilon\mu_1\gamma^2 \\ \lim\limits_{N\to\infty} \frac{\langle \hat{\boldsymbol{\beta}}^0, \boldsymbol{\beta}_0 \rangle}{n} & \lim\limits_{N\to\infty} \frac{\|\hat{\boldsymbol{\beta}}^0\|_2^2}{n} & _\epsilon\mu_1 \lim\limits_{N\to\infty} \frac{\langle \hat{\boldsymbol{\beta}}^0, \boldsymbol{\beta}_0 \rangle}{n} \\ _\epsilon\mu_1\gamma^2 & _\epsilon\mu_1 \lim\limits_{N\to\infty} \frac{\langle \hat{\boldsymbol{\beta}}^0, \boldsymbol{\beta}_0 \rangle}{n} & _\epsilon\tau_1^2\kappa + {}_\epsilon\mu_1^2\gamma^2 \end{bmatrix} .
$$

Calculating the nonzero entries of $_\epsilon\boldsymbol{\Sigma}^4$ yields that

$$\begin{bmatrix} _\epsilon\boldsymbol{\Sigma}^4_{2,2} & _\epsilon\boldsymbol{\Sigma}^4_{2,4} \\ _\epsilon\boldsymbol{\Sigma}^4_{4,2} & _\epsilon\boldsymbol{\Sigma}^4_{4,4} \end{bmatrix} = \begin{bmatrix} _\epsilon\tau_1^2 & _\epsilon\rho_{1,2} \\ _\epsilon\rho_{1,2} & _\epsilon\tau_2^2 \end{bmatrix}, \tag{S45}$$

where

$$_\epsilon\tau_2^2 = \kappa^{-2}\mathrm{E}\left[\{\Psi^1(h_\epsilon(\boldsymbol{Z}_1^3,\bar{\varepsilon}),\boldsymbol{Z}_2^3)\}^2\right]$$
$$_\epsilon\rho_{1,2} = \kappa^{-2}\mathrm{E}\left[\Psi^0(h_\epsilon(\boldsymbol{Z}_1^3,\bar{\varepsilon}),\boldsymbol{Z}_2^3)\Psi^1(h_\epsilon(\boldsymbol{Z}_1^3,\bar{\varepsilon}),\boldsymbol{Z}_3^3)\right] .$$

Consequently, almost surely

$$\lim_{N\to\infty}\frac{1}{p}\sum_{j=1}^{p}\psi\left(\begin{bmatrix} _\epsilon\boldsymbol{\nu}_j^1 \\ _\epsilon\boldsymbol{\nu}_j^2 \end{bmatrix},\begin{bmatrix} \boldsymbol{\beta}_{0,j} \\ 0 \end{bmatrix}\right) = \mathrm{E}\left[\psi\left(\boldsymbol{Z}^4,\begin{bmatrix} \bar{\beta} \\ 0 \end{bmatrix}\right)\right],$$

where $\boldsymbol{Z}^4 = [\boldsymbol{Z}_1^4,\boldsymbol{Z}_2^4]^\top$ is bivariate normal with mean zero and covariance matrix (S45). Iterating further,

$$\begin{bmatrix} _\epsilon\boldsymbol{\Sigma}^5_{1,1} & _\epsilon\boldsymbol{\Sigma}^5_{1,3} & _\epsilon\boldsymbol{\Sigma}^5_{1,5} & _\epsilon\boldsymbol{\Sigma}^5_{1,7} \\ _\epsilon\boldsymbol{\Sigma}^5_{3,1} & _\epsilon\boldsymbol{\Sigma}^5_{3,3} & _\epsilon\boldsymbol{\Sigma}^5_{3,5} & _\epsilon\boldsymbol{\Sigma}^5_{3,7} \\ _\epsilon\boldsymbol{\Sigma}^5_{5,1} & _\epsilon\boldsymbol{\Sigma}^5_{5,3} & _\epsilon\boldsymbol{\Sigma}^5_{5,5} & _\epsilon\boldsymbol{\Sigma}^5_{5,7} \\ _\epsilon\boldsymbol{\Sigma}^5_{7,1} & _\epsilon\boldsymbol{\Sigma}^5_{7,3} & _\epsilon\boldsymbol{\Sigma}^5_{7,5} & _\epsilon\boldsymbol{\Sigma}^5_{7,7} \end{bmatrix} = \begin{bmatrix} \lim\limits_{N\to\infty}\frac{\|\boldsymbol{\beta}_0\|_2^2}{n} & \lim\limits_{N\to\infty}\frac{\langle\hat{\boldsymbol{\beta}}^0,\boldsymbol{\beta}_0\rangle}{n} & _\epsilon\mu_1\gamma^2 & _\epsilon\mu_2\gamma^2 \\ \lim\limits_{N\to\infty}\frac{\langle\hat{\boldsymbol{\beta}}^0,\boldsymbol{\beta}_0\rangle}{n} & \lim\limits_{N\to\infty}\frac{\|\hat{\boldsymbol{\beta}}^0\|_2^2}{n} & _\epsilon\mu_1\lim\limits_{N\to\infty}\frac{\langle\hat{\boldsymbol{\beta}}^0,\boldsymbol{\beta}_0\rangle}{n} & _\epsilon\mu_2\lim\limits_{N\to\infty}\frac{\langle\hat{\boldsymbol{\beta}}^0,\boldsymbol{\beta}_0\rangle}{n} \\ _\epsilon\mu_1\gamma^2 & _\epsilon\mu_1\lim\limits_{N\to\infty}\frac{\langle\hat{\boldsymbol{\beta}}^0,\boldsymbol{\beta}_0\rangle}{n} & _\epsilon\tau_1^2\kappa+_\epsilon\mu_1^2\gamma^2 & _\epsilon\rho_{1,2}\kappa+_\epsilon\mu_1{}_\epsilon\mu_2\gamma^2 \\ _\epsilon\mu_2\gamma^2 & _\epsilon\mu_2\lim\limits_{N\to\infty}\frac{\langle\hat{\boldsymbol{\beta}}^0,\boldsymbol{\beta}_0\rangle}{n} & _\epsilon\rho_{1,2}\kappa+_\epsilon\mu_1{}_\epsilon\mu_2\gamma^2 & _\epsilon\tau_2^2\kappa+_\epsilon\mu_2^2\gamma^2 \end{bmatrix}, \tag{S46}$$

so that almost surely

$$\lim_{N\to\infty}\frac{1}{n}\sum_{j=1}^{n}\psi\left(\begin{bmatrix} \boldsymbol{\eta}_{0,j} \\ _\epsilon\boldsymbol{R}_j^0 \\ _\epsilon\boldsymbol{R}_j^1 \\ _\epsilon\boldsymbol{R}_j^2 \end{bmatrix},\begin{bmatrix} \bar{\varepsilon}_i \\ 0 \\ 0 \\ 0 \end{bmatrix}\right) = \mathrm{E}\left[\psi\left(\boldsymbol{Z}^5,\begin{bmatrix} \bar{\varepsilon} \\ 0 \\ 0 \\ 0 \end{bmatrix}\right)\right],$$

where once again $\boldsymbol{Z}^5 = [\boldsymbol{Z}_1^5,\ldots,\boldsymbol{Z}_4^5]^\top \perp \bar{\varepsilon}$ is multivariate normal with mean zero and covariance matrix (S46). Hence continuing,

$$\begin{bmatrix} _\epsilon\boldsymbol{\Sigma}^6_{2,2} & _\epsilon\boldsymbol{\Sigma}^6_{2,4} & _\epsilon\boldsymbol{\Sigma}^6_{2,6} \\ _\epsilon\boldsymbol{\Sigma}^6_{4,2} & _\epsilon\boldsymbol{\Sigma}^6_{4,4} & _\epsilon\boldsymbol{\Sigma}^6_{4,6} \\ _\epsilon\boldsymbol{\Sigma}^6_{6,2} & _\epsilon\boldsymbol{\Sigma}^6_{6,4} & _\epsilon\boldsymbol{\Sigma}^6_{6,6} \end{bmatrix} = \begin{bmatrix} _\epsilon\tau_1^2 & _\epsilon\rho_{1,2} & _\epsilon\rho_{1,3} \\ _\epsilon\rho_{1,2} & _\epsilon\tau_2^2 & _\epsilon\rho_{2,3} \\ _\epsilon\rho_{1,3} & _\epsilon\rho_{2,3} & _\epsilon\tau_3^2 \end{bmatrix}, \tag{S47}$$

where

$$_\epsilon\tau_3^2 = \kappa^{-2}\mathrm{E}\left[\{\Psi^2(h_\epsilon(\boldsymbol{Z}_1^5,\bar{\varepsilon}),\boldsymbol{Z}_2^5)\}^2\right]$$
$$_\epsilon\rho_{l,m} = \kappa^{-2}\mathrm{E}\left[\Psi^{l-1}(h_\epsilon(\boldsymbol{Z}_1^5,\bar{\varepsilon}),\boldsymbol{Z}_{l+1}^5)\Psi^{m-1}(h_\epsilon(\boldsymbol{Z}_1^5,\bar{\varepsilon}),\boldsymbol{Z}_{m+1}^5)\right] .$$

Therefore almost surely,

$$\lim_{N\to\infty}\frac{1}{p}\sum_{j=1}^{p}\psi\left(\begin{bmatrix} _\epsilon\boldsymbol{\nu}_j^1 \\ _\epsilon\boldsymbol{\nu}_j^2 \\ _\epsilon\boldsymbol{\nu}_j^3 \end{bmatrix},\begin{bmatrix} \boldsymbol{\beta}_{0,j} \\ 0 \\ 0 \end{bmatrix}\right) = \mathrm{E}\left[\psi\left(\boldsymbol{Z}^6,\begin{bmatrix} \bar{\beta} \\ 0 \\ 0 \end{bmatrix}\right)\right],$$

with $\boldsymbol{Z}^6 = [\boldsymbol{Z}_1^6,\boldsymbol{Z}_2^6,\boldsymbol{Z}_3^6]^\top$ trivariate normal with mean zero and covariance (S47). Continuing in this

vein, one gets that for all $t \in [q]$, by Lemmas B.2 and Theorem B.1, almost surely

$$\lim_{N \to \infty} \frac{1}{p} \sum_{j=1}^{p} \psi \left({}_\epsilon \boldsymbol{\nu}_j^t, \boldsymbol{\beta}_{0,j} \right) = \mathrm{E} \left[\psi(\tau_t G, \bar{\beta}) \right]$$

$$\lim_{N \to \infty} \frac{1}{n} \sum_{j=1}^{n} \psi \left(\begin{bmatrix} \boldsymbol{\eta}_{0,j} \\ {}_\epsilon \boldsymbol{R}_j^t \end{bmatrix}, \begin{bmatrix} \bar{\varepsilon}_i \\ 0 \end{bmatrix} \right) = \mathrm{E} \left[\psi \left(\begin{bmatrix} \boldsymbol{Z}_t^{2t+1} \\ \boldsymbol{Z}_{t+2}^{2t+1} \end{bmatrix}, \begin{bmatrix} \bar{\varepsilon} \\ 0 \end{bmatrix} \right) \right] ,$$
(S48)

where $G \sim \mathcal{N}(0,1) \perp \bar{\beta}$, and $(\boldsymbol{Z}_t^{2t+1}, \boldsymbol{Z}_{t+2}^{2t+1}) \perp \bar{\varepsilon}$ are bivariate normal with mean zero and covariance

$$\begin{bmatrix} {}_\epsilon \boldsymbol{\Sigma}_{1,1}^{t+2} & {}_\epsilon \boldsymbol{\Sigma}_{1,t+2}^{t+2} \\ {}_\epsilon \boldsymbol{\Sigma}_{t+2,1}^{t+2} & {}_\epsilon \boldsymbol{\Sigma}_{t+2,t+2}^{t+2} \end{bmatrix} = \begin{bmatrix} \lim_{N \to \infty} \frac{\|\boldsymbol{\beta}_0\|_2^2}{n} & {}_\epsilon \mu_t \gamma^2 \\ {}_\epsilon \mu_t \gamma^2 & {}_\epsilon \tau_t^2 \kappa + {}_\epsilon \mu_t^2 \gamma^2 \end{bmatrix} ,$$
(S49)

and

$${}_\epsilon \tau_t = \kappa^{-2} \mathrm{E} \left[\left\{ \Psi^{t-1}(h_\epsilon(\boldsymbol{Z}_1^{2t-1}, \bar{\varepsilon}), \boldsymbol{Z}_{t+1}^{2t-1}) \right\}^2 \right] ,$$
(S50)

and where $\boldsymbol{Z}_1^{2t-1}, \boldsymbol{Z}_{t+1}^{2t-1} \perp \bar{\varepsilon}$ are bivariate normal with the covariance structure of (S49) with ${}_\epsilon \mu_{t-1}, {}_\epsilon \tau_{t-1}$ in place of ${}_\epsilon \mu_t, {}_\epsilon \tau_t$ respectively.

Next, note that with initialisations ${}_\epsilon \mu_0 = \gamma^{-2} \lim_{n \to \infty} \langle \hat{\boldsymbol{\beta}}^0, \boldsymbol{\beta}_0 \rangle / n$ and ${}_\epsilon \sigma_0^2 = \lim_{n \to \infty} \|\hat{\boldsymbol{\beta}}^0 - \mu_0 \boldsymbol{\beta}_0\|_2^2 / p$, the variances of (S44) correspond to

$$\begin{bmatrix} \gamma^2 & {}_\epsilon \mu_0 \gamma^2 \\ {}_\epsilon \mu_0 \gamma^2 & {}_\epsilon \sigma_0^2 \kappa + {}_\epsilon \mu_0^2 \gamma^2 {}_\epsilon \mu_0 \gamma^2 \end{bmatrix} ,$$

Hence by the definition of ${}_\epsilon \sigma_t$ in (S38), ${}_\epsilon \tau_1 = {}_\epsilon \sigma_1$. By induction it follows that ${}_\epsilon \tau_t = {}_\epsilon \sigma_t$ for all t under consideration. This concludes the proof. $\qquad \square$

B.4.5 Asymptotic equivalence of target and centred AMP

Next it is possible to establish the asymptotic equivalence of the centred recursion (S12)-(S15) with the initial target recursion (S10)-(S11) for appropriately centred iterates. The proof closely follows along the proof of Lemma 4 of the supplementary material of Sur and Candès (2019) which in turn is an adaptation of the proof of Lemma 6.7 of Donoho and Montanari (2016).

Lemma B.4. *In the setting of Corollary B.2.1, it holds that for initial conditions ${}_\epsilon \boldsymbol{\nu}^0 = \hat{\boldsymbol{\beta}}^0 - \mu_0 \boldsymbol{\beta}_0$, ${}_\epsilon \boldsymbol{R}^0 = {}_\epsilon \boldsymbol{S}^0$, $t \in [q]$,*

$$\lim_{N \to \infty} \frac{1}{p} \left\| {}_\epsilon \boldsymbol{\nu}^t - {}_\epsilon \hat{\boldsymbol{\beta}}^t - {}_\epsilon \mu_t \boldsymbol{\beta}_0 \right\|_2^2 = 0, \quad \lim_{N \to \infty} \frac{1}{n} \left\| {}_\epsilon \boldsymbol{R}^t - {}_\epsilon \boldsymbol{S}^t \right\|_2^2 = 0 ,$$

almost surely.

Proof. To facilitate notation, let ${}_\epsilon \boldsymbol{\theta}^t = {}_\epsilon \hat{\boldsymbol{\beta}}^t - {}_\epsilon \mu_t \boldsymbol{\beta}_0$. Then by construction of recursions (S10)-(S11) and (S12)-(S13), and the triangle inequality it holds that

$$\left\| {}_\epsilon \boldsymbol{R}^t - {}_\epsilon \boldsymbol{S}^t \right\|_2 \leq \left\| \boldsymbol{X} \right\|_2 \left\| {}_\epsilon \boldsymbol{\nu}^t - {}_\epsilon \boldsymbol{\theta}^t \right\|_2 + \left\| \Psi^{t-1}({}_\epsilon \boldsymbol{y}^*, {}_\epsilon \boldsymbol{R}^{t-1}) - \Psi^{t-1}({}_\epsilon \boldsymbol{y}^*, {}_\epsilon \boldsymbol{S}^{t-1}) \right\|$$
$$\leq \left\| \boldsymbol{X} \right\|_2 \left\| {}_\epsilon \boldsymbol{\nu}^t - {}_\epsilon \boldsymbol{\theta}^t \right\|_2 + \left\| {}_\epsilon \boldsymbol{R}^{t-1} - {}_\epsilon \boldsymbol{S}^{t-1} \right\|_2 ,$$
(S51)

where the second inequality follows by Lemma G.2 iii) and where ${}_\epsilon \boldsymbol{y}^* = \alpha h_\epsilon(\boldsymbol{\eta}_0, \bar{\varepsilon}) + (1 - \alpha)/2$.

Similarly,

$$\left\|_\epsilon\boldsymbol{\nu}^t - {}_\epsilon\boldsymbol{\theta}^t\right\|_2 \le \left\|_\epsilon\boldsymbol{\nu}^{t-1} - {}_\epsilon\boldsymbol{\theta}^{t-1}\right\|_2 + |_\epsilon q_{t-1} - 1|\left\|_\epsilon\boldsymbol{\nu}^{t-1} + {}_\epsilon\mu_{t-1}\boldsymbol{\beta}_0\right\|_2 + \left\|_\epsilon\mu_t - {}_\epsilon a_t\right\|\|\boldsymbol{\beta}_0\| + \frac{1}{\kappa}\|\boldsymbol{X}\|_2\left\|_\epsilon\boldsymbol{R}^{t-1} - {}_\epsilon\boldsymbol{S}^{t-1}\right\|_2 .$$

Now since $_\epsilon\boldsymbol{\nu}^0 = {}_\epsilon\boldsymbol{\theta}^0$, an induction argument on the number of iterations t gives that there exists a constant C, which may depend on κ, such that for each $t \in [q]$,

$$\left\|_\epsilon\boldsymbol{\nu}^t - {}_\epsilon\boldsymbol{\theta}^t\right\|_2 \le (C\|\boldsymbol{X}\|_2)^{2t}\left(\sum_{l=1}^{t-1}|_\epsilon q_t - 1|\left\|_\epsilon\boldsymbol{\nu}^l + {}_\epsilon\mu_l\boldsymbol{\beta}_0\right\|_2 + \sum_{l=0}^{t-1}|_\epsilon\mu_l - a_l|\|\boldsymbol{\beta}_0\|_2\right). \tag{S52}$$

Recalling that $_\epsilon q_t = -\frac{1}{\kappa n}\sum_{j=1}^n \frac{-b_t\zeta''(\mathrm{prox}_{b_t\zeta}(b_t h_\epsilon(\boldsymbol{\eta}_{0,j}, \bar\varepsilon_j), \boldsymbol{R}_j^t))}{1 + b_t\zeta''(\mathrm{prox}_{b_t\zeta}(b_t h_\epsilon(\boldsymbol{\eta}_{0,j}, \bar\varepsilon_j), \boldsymbol{R}_j^t))}$ (cf. (S14) and (S90)), it follows from Lemma G.2 iv) and Corollary B.2.1 i), that almost surely

$$\lim_{N\to\infty} {}_\epsilon q_t = \frac{1}{\kappa}\mathrm{E}\left[\frac{b_t\zeta''(\mathrm{prox}_{b_t\zeta}(b_t h_\epsilon(Z_1^t, \bar\varepsilon), {}_\epsilon Z_2^t))}{1 + b_t\zeta''(\mathrm{prox}_{b_t\zeta}(b_t h_\epsilon(Z_1^t, \bar\varepsilon), {}_\epsilon Z_2^t))}\right] \tag{S53}$$

for $Z_1^t \sim N(0, \gamma^2)$, $_\epsilon Z_2^t = {}_\epsilon\mu_t Z_1^t + \kappa^{1/2}{}_\epsilon\sigma_t G$, $G \sim N(0,1)$, $\bar\varepsilon \sim U(0,1)$ and where $G \perp \bar\beta$, $G \perp Z_1^t$ and $(Z_1^t, {}_\epsilon Z_2^t) \perp \bar\varepsilon$. Now the right hand side of (S53) can be rearranged as

$$\frac{1}{\kappa}\mathrm{E}\left[\frac{b_t\zeta''(\mathrm{prox}_{b_t\zeta}(b_t h_\epsilon(Z_1^t, \bar\varepsilon), {}_\epsilon Z_2^t))}{1 + b_t\zeta''(\mathrm{prox}_{b_t\zeta}(b_t h_\epsilon(Z_1^t, \bar\varepsilon), {}_\epsilon Z_2^t))}\right] = \frac{1}{\kappa}\mathrm{E}\left[1 - \frac{1}{b_t\zeta''(\mathrm{prox}_{b_t\zeta}(b_t h_\epsilon(Z_1^t, \bar\varepsilon), {}_\epsilon Z_2^t))}\right] = 1 ,$$

according to state evolution equation (S36).

Again, by Corollary B.2.1 i) and $\psi(u, v) = u^2$, it follows that almost surely,

$$\lim_{N\to\infty}\sum_{j=1}(_\epsilon\boldsymbol{\nu}_j^t)^2 = \mathrm{E}\left[(_\epsilon\sigma_t Z)^2\right] < \infty ,$$

for $Z \sim N(0,1)$. Further recall that by assumption $\lim_{n\to\infty}\|\boldsymbol{\beta}_0\|_2^2/n = \gamma^2$ and that by Lemma G.3, $\lim_{N\to\infty}\|\boldsymbol{X}\|_2 < \infty$ almost surely. Hence it holds almost surely that

$$\lim_{N\to\infty}\frac{1}{\sqrt{p}}(C\|\boldsymbol{X}\|_2)^{2t}\sum_{l=0}^{t-1}|_\epsilon q_l - 1|\left\|_\epsilon\boldsymbol{\nu}_j^l + {}_\epsilon\mu_l\boldsymbol{\beta}_0\right\|_2 = 0 . \tag{S54}$$

Similarly, recalling the definition of $_\epsilon a_t$ in (S41) and invoking Lemma G.2 iv) and Corollary B.2.1 equation (S48), almost surely

$$\lim_{N\to\infty} {}_\epsilon a_t = \frac{1}{\kappa}\mathrm{E}\left[\left.\frac{\partial}{\partial Z}\Psi^{t-1}(h_\epsilon(Z, \bar\varepsilon), {}_\epsilon Z_2^t)\right|_{Z=Z_1^t}\right], \tag{S55}$$

for $Z_1^t \sim N(0, \gamma^2)$, $_\epsilon Z_2^t = {}_\epsilon\mu_t Z_1^t + \kappa^{1/2}{}_\epsilon\sigma_t G$, $G \sim N(0,1)$, $\bar\varepsilon \sim U(0,1)$ and where $G \perp \bar\beta$, $G \perp Z_1^t$ and $(Z_1^t, {}_\epsilon Z_2^t) \perp \bar\varepsilon$.

To compute the expectation in (S55), Stein's Lemma (Stein, 1981, Lemma 1) shall prove useful. It states that for a normal random variable $z \sim N(\mu, \sigma^2)$, and a differentiable function $h(z)$, $\mathrm{E}[h(z)(z - \mu)] = \sigma^2\mathrm{E}(h'(z))$, where h is derivative of h with respect to z.

To use the Lemma, first note that $_\epsilon Z_2^t$ is a function of Z_1^t. Hence write $_\epsilon Z_2^t = t_\epsilon(Z_1^t, G_t) = _\epsilon \mu_t Z_1^t + \kappa^{1/2} {}_\epsilon \sigma_t G_t$. Then conditioning on both $\bar\epsilon, G_t$, one gets that

$$\mathrm{E}\left[\frac{\partial}{\partial Z}\Psi^{t-1}(h_\epsilon(Z,\bar\epsilon), {}_\epsilon Z_2^t)\bigg|_{Z=Z_1^t}\right] = \mathrm{E}_{G_t,\bar\epsilon}\left\{\mathrm{E}_{Z_1^t}\left[\frac{\partial}{\partial Z}\Psi(h_\epsilon(Z,\bar\epsilon), s)\bigg|_{\substack{Z=Z_1^t,\\ s=t_\epsilon(Z_1^t,G_t)}}\bigg|G_t,\bar\epsilon\right]\right\}. \quad \text{(S56)}$$

Now since Z_1^t is independent of $G_t, \bar\epsilon$, the conditional distribution of Z_1^t given $G_t, \bar\epsilon$ is its unconditional distribution, i.e. $Z_1^t|_{G_t,\bar\epsilon} \sim N(0,\gamma^2)$ and thus by Stein's Lemma,

$$\gamma^2 \mathrm{E}_{Z_1^t}\left[\frac{\partial}{\partial Z_1^t}\Psi(h_\epsilon(Z_1^t,\bar\epsilon), t_\epsilon(Z_1^t,G_t))\bigg|G_t,\bar\epsilon\right] = \mathrm{E}_{Z_1^t}\left[Z_1^t\Psi(h_\epsilon(Z_1^t,\bar\epsilon), t_\epsilon(Z_1^t,G_t))\bigg|G_t,\bar\epsilon\right]. \quad \text{(S57)}$$

Now an application of the chain rule to the left hand side of (S57) yields

$$\mathrm{E}_{Z_1^t}\left[\frac{\partial}{\partial Z_1^t}\Psi(h_\epsilon(Z_1^t,\bar\epsilon), t_\epsilon(Z_1^t,G_t))\bigg|G_t,\bar\epsilon\right] = \mathrm{E}_{Z_1^t}\left[\frac{\partial}{\partial Z}\Psi(h_\epsilon(Z,\bar\epsilon), s)\bigg|_{\substack{Z=Z_1^t,\\ s=t_\epsilon(Z_1^t,G_t)}}\bigg|G_t,\bar\epsilon\right]$$
$$+ {}_\epsilon \mu_{t-1}\mathrm{E}_{Z_1^t}\left[\frac{\partial}{\partial s}\Psi(h_\epsilon(Z,\bar\epsilon), s)\bigg|_{\substack{Z=Z_1^t,\\ s=t_\epsilon(Z_1^t,G_t)}}\bigg|G_t,\bar\epsilon\right] \quad \text{(S58)}$$

Combining (S57) and (S58) gives

$$\mathrm{E}_{Z_1^t}\left[\frac{\partial}{\partial Z}\Psi(h_\epsilon(Z,\bar\epsilon), s)\bigg|_{\substack{Z=Z_1^t,\\ s=t_\epsilon(Z_1^t,G_t)}}\bigg|G_t,\bar\epsilon\right] = \frac{1}{\gamma^2}\mathrm{E}_{Z_1^t}\left[Z_1^t\Psi(h_\epsilon(Z_1^t,\bar\epsilon), t_\epsilon(Z_1^t,G_t))\bigg|G_t,\bar\epsilon\right]$$
$$- {}_\epsilon \mu_{t-1}\mathrm{E}_{Z_1^t}\left[\frac{\partial}{\partial s}\Psi(h_\epsilon(Z,\bar\epsilon), s)\bigg|_{\substack{Z=Z_1^t,\\ s=t_\epsilon(Z_1^t,G_t)}}\bigg|G_t,\bar\epsilon\right]$$

Substituting back onto (S56) in conjunction with (S55) gives that almost surely,

$$\lim_{N\to\infty} {}_\epsilon a_t = \frac{1}{\kappa\gamma^2}\mathrm{E}\left[Z_1^t\Psi(h_\epsilon(Z_1^t,\bar\epsilon), {}_\epsilon Z_2^t)\right] - \frac{_\epsilon \mu_{t-1}}{\kappa}\mathrm{E}\left[\frac{\partial}{\partial s}\Psi(h_\epsilon(Z,\bar\epsilon), s)\bigg|_{\substack{Z=Z_1^t,\\ s={}_\epsilon Z_2^t}}\right].$$

Now using (S36) and (S90), it follows that

$$-\frac{1}{\kappa}\mathrm{E}\left[\frac{\partial}{\partial s}\Psi(h_\epsilon(Z,\bar\epsilon), s)\bigg|_{\substack{Z=Z_1^t,\\ s={}_\epsilon Z_2^t}}\right] = \frac{1}{\kappa}\mathrm{E}\left[\frac{b\zeta''(\mathrm{prox}_{b\zeta}(bh_\epsilon(Z,\bar\epsilon) + {}_\epsilon Z_2^t))}{1 + b\zeta''(\mathrm{prox}_{b\zeta}(bh_\epsilon(Z,\bar\epsilon + {}_\epsilon Z_2^t)))}\right]$$
$$= \frac{1}{\kappa}\mathrm{E}\left[1 - \frac{1}{1 + b\zeta''(\mathrm{prox}_{b\zeta}(bh_\epsilon(Z,\bar\epsilon + {}_\epsilon Z_2^t)))}\right]$$
$$= 1.$$

25

Hence, the state evolution of $_\epsilon\mu_t$ as specified in (S37) gives that almost surely

$$
\begin{aligned}
\lim_{N\to\infty} {}_\epsilon a_t &= \frac{1}{\kappa\gamma^2}\mathrm{E}\left[Z_1^t\Psi(h_\epsilon(Z_1^t,\bar\varepsilon),{}_\epsilon Z_2^t)\right] - \frac{\epsilon\mu_{t-1}}{\kappa}\mathrm{E}\left[\left.\frac{\partial}{\partial s}\Psi(h_\epsilon(Z,\bar\varepsilon),s)\right|_{\substack{Z=Z_1^t,\\ s={}_\epsilon Z_2^t}}\right] \\
&= \frac{1}{\kappa\gamma^2}\mathrm{E}\left[Z_1^t\Psi(h_\epsilon(Z_1^t,\bar\varepsilon),{}_\epsilon Z_2^t)\right] + {}_\epsilon\mu_{t-1} \\
&= {}_\epsilon\mu_t\,.
\end{aligned}
\tag{S59}
$$

Hence, for any $l\in\mathbb{N}_0$, $\lim\limits_{N\to\infty}|_\epsilon a_t - {}_\epsilon\mu_t| = 0$ almost surely. By almost sure boundedness of $\boldsymbol{X}$ and since $\|\boldsymbol{\beta}_0\|_2^2/n \to \gamma^2$ as $n\to\infty$, it follows that almost surely

$$
\lim_{N\to\infty}\frac{1}{\sqrt{p}}\sum_{l=0}^{t-1}|_\epsilon a_t - {}_\epsilon\mu_t|\,\|\boldsymbol{\beta}_0\|_2 = 0\,.
\tag{S60}
$$

In conjunction with (S52) and (S54) and (S60), this yields that almost surely,

$$
\left\|{}_\epsilon\boldsymbol{\nu}^t - {}_\epsilon\boldsymbol{\theta}^t\right\|_2 = 0\,.
\tag{S61}
$$

Almost sure convergence of $\|{}_\epsilon\boldsymbol{R}^t - {}_\epsilon\boldsymbol{S}^t\|_2$ then follows by an induction argument on t, (S51), (S61) and since $_\epsilon\boldsymbol{S}^0 = {}_\epsilon\boldsymbol{R}^0$. This concludes the proof. □

B.4.6 Asymptotic equivalence of AMP and MDYPL

Theorem B.3. *In the setting of Lemma B.4 and assuming that there exists a stationary solution $(_\epsilon\mu_*, {}_\epsilon b_*, {}_\epsilon\sigma_*)$ to the state evolution (S37)-(S38). Then, if $\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\|_2/\sqrt{n} = \mathcal{O}(1)$ almost surely,*

$$
\lim_{\epsilon\to0}\lim_{t\to\infty}\lim_{N\to\infty}\frac{1}{p}\left\|{}_\epsilon\hat{\boldsymbol{\beta}}^t - \hat{\boldsymbol{\beta}}^{\mathrm{DY}}\right\|_2^2 = 0
$$

almost surely.

Proof. First, a second order Taylor approximation of $\ell(\hat{\boldsymbol{\beta}}^{\mathrm{DY}})$ around $_\epsilon\hat{\boldsymbol{\beta}}^t$ gives

$$
\ell(\hat{\boldsymbol{\beta}}^{\mathrm{DY}}) = \ell(_\epsilon\hat{\boldsymbol{\beta}}^t) + \nabla\ell(_\epsilon\hat{\boldsymbol{\beta}}^t)^\top(\hat{\boldsymbol{\beta}}^{\mathrm{DY}} - {}_\epsilon\hat{\boldsymbol{\beta}}^t) + \frac{1}{2}\boldsymbol{r}\,,
\tag{S62}
$$

where the remainder term $\boldsymbol{r}$ is, equal to

$$
\boldsymbol{r} = (\hat{\boldsymbol{\beta}}^{\mathrm{DY}} - {}_\epsilon\hat{\boldsymbol{\beta}}^t)^\top\nabla\nabla^\top\ell(\hat{\boldsymbol{\beta}}^{\mathrm{DY}} - c(_\epsilon\hat{\boldsymbol{\beta}}^t - \hat{\boldsymbol{\beta}}^{\mathrm{DY}}))(\hat{\boldsymbol{\beta}}^{\mathrm{DY}} - {}_\epsilon\hat{\boldsymbol{\beta}}^t)\,,
$$

for some constant $c\in[0,1]$. Now by Sur 2019, Lemma 7, with probability at least $1 - c_1\exp\{-c_2 n\}$ for absolute constants $c_1, c_2 > 0$, it holds that for all $\boldsymbol{\beta}\in\Re^p$

$$
\nabla\nabla^\top\ell(\boldsymbol{\beta}) \succ \omega\left(\frac{\|\boldsymbol{\beta}\|_2}{\sqrt{n}}\right)\boldsymbol{I}_p\,,
$$

where $\boldsymbol{A}\succ\boldsymbol{B}$ means that $\boldsymbol{A}-\boldsymbol{B}$ is positive definite. Here, $\omega(\cdot)\in(0,1)$ is a nonincreasing function independent of n. Clearly,

$$
\left\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}} - c(\hat{\boldsymbol{\beta}}^{\mathrm{DY}} - {}_\epsilon\hat{\boldsymbol{\beta}}^t)\right\|_2 \leq \max\left\{\left\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\right\|_2, \left\|{}_\epsilon\hat{\boldsymbol{\beta}}^t\right\|_2\right\}\,,
$$

so that with probability at least $1 - c_1 \exp\{-c_2 n\}$

$$\boldsymbol{r} \geq \left\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}} - {}_\epsilon\hat{\boldsymbol{\beta}}^t\right\|_2^2 \omega\left(\frac{\left\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}} - c(\hat{\boldsymbol{\beta}}^{\mathrm{DY}} - {}_\epsilon\hat{\boldsymbol{\beta}}^t)\right\|_2}{\sqrt{n}}\right)$$

$$\geq \left\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}} - {}_\epsilon\hat{\boldsymbol{\beta}}^t\right\|_2^2 \omega\left(\max\left\{\frac{\left\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\right\|_2}{\sqrt{n}}, \frac{\left\|{}_\epsilon\hat{\boldsymbol{\beta}}^t\right\|_2}{\sqrt{n}}\right\}\right). \tag{S63}$$

Using (S63) and the Cauchy-Schwarz inequality, (S62) can be rearranged as

$$\left\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}} - {}_\epsilon\hat{\boldsymbol{\beta}}^t\right\|_2 \leq \frac{2}{\omega\left(\max\left\{\frac{\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\|_2}{\sqrt{n}}, \frac{\|{}_\epsilon\hat{\boldsymbol{\beta}}^t\|_2}{\sqrt{n}}\right\}\right)} \left\|\nabla\ell({}_\epsilon\hat{\boldsymbol{\beta}}^t)\right\|_2,$$

so that it remains to control $\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\|_2, \|{}_\epsilon\hat{\boldsymbol{\beta}}^t\|_2$ and $\|\nabla\ell({}_\epsilon\hat{\boldsymbol{\beta}}^t)\|_2$. Towards bounding $\|\nabla\ell({}_\epsilon\hat{\boldsymbol{\beta}}^t)\|_2$, recall that by (S34) and (S35), any by the definition of the function $\Psi^t(\cdot,\cdot)$, and the identity $z - \zeta'(\mathrm{prox}_{b\zeta}(z)) = \mathrm{prox}_{b\zeta}(z)$,

$$-\mathrm{prox}_{\epsilon b_t \zeta}\left({}_\epsilon b_t {}_\epsilon\boldsymbol{y}^* + {}_\epsilon\boldsymbol{S}^{t-1}\right) = {}_\epsilon\boldsymbol{S}^t - {}_\epsilon\boldsymbol{S}^{t-1} + \boldsymbol{X}_\epsilon\hat{\boldsymbol{\beta}}^t,$$

and thus

$${}_\epsilon\hat{\boldsymbol{\beta}}^t - {}_\epsilon\hat{\boldsymbol{\beta}}^{t-1} = \frac{\epsilon b_t}{\kappa}\boldsymbol{X}^\top\left\{{}_\epsilon\boldsymbol{y}^* - \zeta'\left({}_\epsilon\boldsymbol{S}^t - {}_\epsilon\boldsymbol{S}^{t-1} + \boldsymbol{X}_\epsilon\hat{\boldsymbol{\beta}}^t\right)\right\}.$$

Therefore,

$$\nabla\ell({}_\epsilon\hat{\boldsymbol{\beta}}^t) = \boldsymbol{X}^\top\left\{\boldsymbol{y} - \zeta'(\boldsymbol{X}_\epsilon\hat{\boldsymbol{\beta}}^t)\right\}$$

$$= \boldsymbol{X}^\top\{\boldsymbol{y} - {}_\epsilon\boldsymbol{y}\}$$

$$+ \boldsymbol{X}^\top\left\{{}_\epsilon\boldsymbol{y} - \zeta'\left({}_\epsilon\boldsymbol{S}^t - {}_\epsilon\boldsymbol{S}^{t-1} - \boldsymbol{X}_\epsilon\hat{\boldsymbol{\beta}}^t\right)\right\}$$

$$+ \boldsymbol{X}^\top\left\{\zeta'\left({}_\epsilon\boldsymbol{S}^t - {}_\epsilon\boldsymbol{S}^{t-1} - \boldsymbol{X}_\epsilon\hat{\boldsymbol{\beta}}^t\right) - \zeta'\left(\boldsymbol{X}_\epsilon\hat{\boldsymbol{\beta}}^t\right)\right\}.$$

almost surely. Further, using that $\zeta'(x)$ is Lipschitz with Lipschitz constant $1/4$,

$$\left\|\boldsymbol{X}^\top\left\{\zeta'\left({}_\epsilon\boldsymbol{S}^t - {}_\epsilon\boldsymbol{S}^{t-1} - \boldsymbol{X}_\epsilon\hat{\boldsymbol{\beta}}^t\right) - \zeta'\left(\boldsymbol{X}_\epsilon\hat{\boldsymbol{\beta}}^t\right)\right\}\right\|_2 \leq \|\!\|\boldsymbol{X}\|\!\|_2\left\|\zeta'\left({}_\epsilon\boldsymbol{S}^t - {}_\epsilon\boldsymbol{S}^{t-1} - \boldsymbol{X}_\epsilon\hat{\boldsymbol{\beta}}^t\right) - \zeta'\left(\boldsymbol{X}_\epsilon\hat{\boldsymbol{\beta}}^t\right)\right\|_2$$

$$\leq \frac{1}{4}\|\!\|\boldsymbol{X}\|\!\|_2\left\|{}_\epsilon\boldsymbol{S}^t - {}_\epsilon\boldsymbol{S}^{t-1} + \boldsymbol{X}_\epsilon\hat{\boldsymbol{\beta}}^t - \boldsymbol{X}_\epsilon\hat{\boldsymbol{\beta}}^t\right\|_2$$

$$= \frac{1}{4}\|\!\|\boldsymbol{X}\|\!\|_2\left\|{}_\epsilon\boldsymbol{S}^t - {}_\epsilon\boldsymbol{S}^{t-1}\right\|_2$$

Using (S34), it immediately follows that

$$\left\|\boldsymbol{X}^\top\left\{{}_\epsilon\boldsymbol{y} - \zeta'\left({}_\epsilon\boldsymbol{S}^t - {}_\epsilon\boldsymbol{S}^{t-1} - \boldsymbol{X}_\epsilon\hat{\boldsymbol{\beta}}^t\right)\right\}\right\|_2 = \frac{\kappa}{\epsilon b^t}\left\|{}_\epsilon\hat{\boldsymbol{\beta}}^t - {}_\epsilon\hat{\boldsymbol{\beta}}^{t-1}\right\|_2.$$

Therefore, with probability at least $1 - c_1 \exp\{-c_2 n\}$,

$$\left\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}} - {}_\epsilon\hat{\boldsymbol{\beta}}^t\right\|_2 \leq \frac{2}{\omega\left(\max\left\{\frac{\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\|_2}{\sqrt{n}}, \frac{\|{}_\epsilon\hat{\boldsymbol{\beta}}^t\|_2}{\sqrt{n}}\right\}\right)}\left\{\frac{\kappa}{\epsilon b^t}\left\|{}_\epsilon\hat{\boldsymbol{\beta}}^t - {}_\epsilon\hat{\boldsymbol{\beta}}^{t-1}\right\|_2 + \frac{1}{4}\|\!\|\boldsymbol{X}\|\!\|_2\left\|{}_\epsilon\boldsymbol{S}^t - {}_\epsilon\boldsymbol{S}^{t-1}\right\|_2 + \|\!\|\boldsymbol{X}\|\!\|_2\left\|\boldsymbol{y} - {}_\epsilon\boldsymbol{y}\right\|_2\right\}.$$

To finish the proof initialise the AMP recursion at the stationary solution, i.e. $({}_\epsilon\mu_0, {}_\epsilon b_0, {}_\epsilon\sigma_0) = ({}_\epsilon\mu_*, {}_\epsilon b_*, {}_\epsilon\sigma_*)$. Then it follows immediately from (S48) with $\psi(x, y) = x^2$, in conjunction with

Lemma B.4 and the stationarity condition that almost surely

$$\lim_{N\to\infty}\frac{\left\|_\epsilon\hat{\boldsymbol{\beta}}^t\right\|_2}{\sqrt{p}}\le\lim_{N\to\infty}\frac{\left\|_\epsilon\hat{\boldsymbol{\beta}}^t-_\epsilon\mu_*\boldsymbol{\beta}_0\right\|_2}{\sqrt{p}}+\lim_{n\to\infty}\frac{\left\|_\epsilon\mu_*\boldsymbol{\beta}_0\right\|_2}{\sqrt{p}}=_\epsilon\sigma_*+\frac{\gamma}{_\epsilon\mu_*\kappa}\,.\tag{S64}$$

Further, $\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\|_2/\sqrt{n}=\mathcal{O}(1)$ almost surely by assumption. Hence, by the Cauchy property of $_\epsilon\boldsymbol{\nu}^t,_\epsilon\boldsymbol{R}^t$, as shown in Lemma G.8 along with the equivalence of recursions of Lemma B.4, (S64) and Lemma G.4 and Lemma G.5, it follows that with probability at least $1-c_1\exp\{-c_n\}$

$$\lim_{\epsilon\to0}\lim_{t\to\infty}\lim_{N\to\infty}\frac{1}{\sqrt{p}}\left\|_\epsilon\hat{\boldsymbol{\beta}}^t-\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\right\|_2=0\,.$$

The almost sure claim comes from the first Borel-Cantelli Lemma (e.g. Shiryaev 2016, p. 308). $\qquad\square$

It is now possible to finish the proof of Theorem 3.1. First of all, note that by Lemma G.6, there exists a unique stationary solution $(_\epsilon\mu_*,_\epsilon b_*,_\epsilon\sigma_*)$ to the state evolution (S37)-(S38) that is continuous in ϵ such that $\lim_{\epsilon\to0}(_\epsilon\mu_*,_\epsilon b_*,_\epsilon\sigma_*)=(\mu_*,b_*,\sigma_*)$, where (μ_*,b_*,σ_*) are a stationary solution to (S16), (S17), (S18).

Now, let ψ be a pseudo Lipschitz function of order 2. Then

$$\left|\frac{1}{p}\sum_{j=1}^p\psi\left(\hat{\boldsymbol{\beta}}_j^{\mathrm{DY}}-\mu_*\boldsymbol{\beta}_{0,j},\boldsymbol{\beta}_{0,j}\right)-\frac{1}{p}\sum_{j=1}^p\psi\left(_\epsilon\boldsymbol{\nu}^t,\boldsymbol{\beta}_{0,j}\right)\right|$$

$$\le\left|\frac{1}{p}\sum_{j=1}^p\psi\left(\hat{\boldsymbol{\beta}}_j^{\mathrm{DY}}-_\epsilon\mu_*\boldsymbol{\beta}_{0,j},\boldsymbol{\beta}_{0,j}\right)-\frac{1}{p}\sum_{j=1}^p\psi\left(_\epsilon\hat{\boldsymbol{\beta}}^t-_\epsilon\mu_*\boldsymbol{\beta}_{0,j},\boldsymbol{\beta}_{0,j}\right)\right|$$

$$+\left|\frac{1}{p}\sum_{j=1}^p\psi\left(_\epsilon\hat{\boldsymbol{\beta}}^t-_\epsilon\mu_*\boldsymbol{\beta}_{0,j},\boldsymbol{\beta}_{0,j}\right)-\frac{1}{p}\sum_{j=1}^p\psi\left(_\epsilon\boldsymbol{\nu}^t,\boldsymbol{\beta}_{0,j}\right)\right|.\tag{S65}$$

Now

$$\sum_{j=1}^p\left|\psi\left(\hat{\boldsymbol{\beta}}_j^{\mathrm{DY}}-_\epsilon\mu_*\boldsymbol{\beta}_{0,j},\boldsymbol{\beta}_{0,j}\right)-\psi\left(_\epsilon\hat{\boldsymbol{\beta}}^t-_\epsilon\mu_*\boldsymbol{\beta}_{0,j},\boldsymbol{\beta}_{0,j}\right)\right|$$

$$\le\sum_{j=1}^p\left|\psi\left(\hat{\boldsymbol{\beta}}_j^{\mathrm{DY}}-_\epsilon\mu_*\boldsymbol{\beta}_{0,j},\boldsymbol{\beta}_{0,j}\right)-\psi\left(_\epsilon\hat{\boldsymbol{\beta}}^t-_\epsilon\mu_*\boldsymbol{\beta}_{0,j},\boldsymbol{\beta}_{0,j}\right)\right|$$

$$\le C\sum_{j=1}^p\left(1+\left\|\begin{bmatrix}\hat{\boldsymbol{\beta}}_j^{\mathrm{DY}}-_\epsilon\mu_*\boldsymbol{\beta}_{0,j}\\\boldsymbol{\beta}_{0,j}\end{bmatrix}\right\|_2+\left\|\begin{bmatrix}_\epsilon\hat{\boldsymbol{\beta}}^t-_\epsilon\mu_*\boldsymbol{\beta}_{0,j}\\\boldsymbol{\beta}_{0,j}\end{bmatrix}\right\|_2\right)\left\|\hat{\boldsymbol{\beta}}_j^{\mathrm{DY}}-_\epsilon\hat{\boldsymbol{\beta}}_j^t\right\|_2$$

$$\le C\sqrt{\sum_{j=1}^p\left(1+\left\|\begin{bmatrix}\hat{\boldsymbol{\beta}}_j^{\mathrm{DY}}-_\epsilon\mu_*\boldsymbol{\beta}_{0,j}\\\boldsymbol{\beta}_{0,j}\end{bmatrix}\right\|_2+\left\|\begin{bmatrix}_\epsilon\hat{\boldsymbol{\beta}}^t-_\epsilon\mu_*\boldsymbol{\beta}_{0,j}\\\boldsymbol{\beta}_{0,j}\end{bmatrix}\right\|_2\right)^2}\left\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}-_\epsilon\hat{\boldsymbol{\beta}}^t\right\|_2,\tag{S66}$$

where the first inequality follows from the triangle inequality, the second by virtue of ψ being pseudo Lipschitz of order 2 and the last by the Cauchy-Schwarz inequality.

Further, letting $\boldsymbol{a}_j=[a_{j1},a_{j2}]^\top=[\hat{\boldsymbol{\beta}}_j^{\mathrm{DY}}-_\epsilon\mu_*\boldsymbol{\beta}_{0,j},\boldsymbol{\beta}_{0,j}]^\top$ and $\boldsymbol{b}_j=[b_{j1},a_{j2}]^\top=[\hat{\boldsymbol{\beta}}_j^t-_\epsilon\mu_*\boldsymbol{\beta}_{0,j},\boldsymbol{\beta}_{0,j}]^\top,$

simple manipulations yield

$$\sum_{j=1}^{p} \left(1 + \|\boldsymbol{a}_j\|_2 + \|\boldsymbol{b}_j\|_2\right)^2 = \sum_{j=1}^{p} 1 + 2\left(\|\boldsymbol{a}_j\|_2 + \|\boldsymbol{b}_j\|_2\right) + \left(\|\boldsymbol{a}_j\|_2 + \|\boldsymbol{b}_j\|_2\right)^2$$

$$= p + 2\sum_{j=1}^{p}\|\boldsymbol{a}_j\|_2 + 2\sum_{j=1}^{p}\|\boldsymbol{b}_j\|_2 + \sum_{j=1}^{p}\|\boldsymbol{a}_j\|_2^2 + \sum_{j=1}^{p}\|\boldsymbol{b}_j\|_2^2 + 2\sum_{j=1}^{p}\|\boldsymbol{a}_j\|_2\|\boldsymbol{b}_j\|_2$$

$$= p + \|\boldsymbol{a}_{\bullet 1}\|_2^2 + \|\boldsymbol{b}_{\bullet 1}\|_2^2 + 2\|\boldsymbol{a}_{\bullet 2}\|_2^2 + 2\sum_{j=1}^{p}\|\boldsymbol{a}_j\|_2 + 2\sum_{j=1}^{p}\|\boldsymbol{b}_j\|_2 + 2\sum_{j=1}^{p}\|\boldsymbol{a}_j\|_2\|\boldsymbol{b}_j\|_2$$

$$\leq p + \|\boldsymbol{a}_{\bullet 1}\|_2^2 + \|\boldsymbol{b}_{\bullet 1}\|_2^2 + 2\|\boldsymbol{a}_{\bullet 2}\|_2^2 + 2\sum_{j=1}^{p}|a_{j1}| + 2\sum_{j=1}^{p}|b_{j1}| + 4\sum_{j=1}^{p}|a_{j2}| + 2\sum_{j=1}^{p}\|\boldsymbol{a}_j\|_2\|\boldsymbol{b}_j\|_2$$

$$= p + \|\boldsymbol{a}_{\bullet 1}\|_2^2 + \|\boldsymbol{b}_{\bullet 1}\|_2^2 + 2\|\boldsymbol{a}_{\bullet 2}\|_2^2 + 2\|\boldsymbol{a}_{\bullet 1}\|_1 + 2\|\boldsymbol{b}_{\bullet 1}\|_1 + 4\|\boldsymbol{a}_{\bullet 2}\|_1 + 2\sum_{j=1}^{p}\|\boldsymbol{a}_j\|_2\|\boldsymbol{b}_j\|_2$$

$$\leq p + \|\boldsymbol{a}_{\bullet 1}\|_2^2 + \|\boldsymbol{b}_{\bullet 1}\|_2^2 + 2\|\boldsymbol{a}_{\bullet 2}\|_2^2 + 2\sqrt{p}\|\boldsymbol{a}_{\bullet 1}\|_2 + 2\sqrt{p}\|\boldsymbol{b}_{\bullet 1}\|_2 + 4\sqrt{p}\|\boldsymbol{a}_{\bullet 2}\|_2 + 2\sum_{j=1}^{p}\|\boldsymbol{a}_j\|_2\|\boldsymbol{b}_j\|_2$$

$$\leq p + \|\boldsymbol{a}_{\bullet 1}\|_2^2 + \|\boldsymbol{b}_{\bullet 1}\|_2^2 + 2\|\boldsymbol{a}_{\bullet 2}\|_2^2 + 2\sqrt{p}\|\boldsymbol{a}_{\bullet 1}\|_2 + 2\sqrt{p}\|\boldsymbol{b}_{\bullet 1}\|_2 + 4\sqrt{p}\|\boldsymbol{a}_{\bullet 2}\|_2$$
$$+ 2\sqrt{\sum_{j=1}^{p} a_{j1}^2 + a_{j2}^2}\sqrt{\sum_{j=1}^{p} b_{j1}^2 + a_{j2}^2}$$

$$\leq p + \|\boldsymbol{a}_{\bullet 1}\|_2^2 + \|\boldsymbol{b}_{\bullet 1}\|_2^2 + 2\|\boldsymbol{a}_{\bullet 2}\|_2^2 + 2\sqrt{p}\|\boldsymbol{a}_{\bullet 1}\|_2 + 2\sqrt{p}\|\boldsymbol{b}_{\bullet 1}\|_2 + 4\sqrt{p}\|\boldsymbol{a}_{\bullet 2}\|_2$$
$$+ 2(\|\boldsymbol{a}_{\bullet 1}\|_2 + \|\boldsymbol{a}_{\bullet 2}\|_2)(\|\boldsymbol{b}_{\bullet 1}\|_2 + \|\boldsymbol{a}_{\bullet 2}\|_2),$$
(S67)

where $\boldsymbol{a}_{\bullet i} = (a_{1i}, \ldots, a_{pi})$, $(i = 1, 2)$ and a similar convention applies to $\boldsymbol{b}_{\bullet 1}$. The fourth line follows since $\sqrt{x^2 + y^2} \leq \sqrt{x^2} + \sqrt{y^2} = |x| + |y|$, the sixth by the Cauchy-Schwarz inequality and since $\|\boldsymbol{v}\|_1 = |\langle \mathbf{1}, |\boldsymbol{v}|\rangle|$, where $\mathbf{1}$ is a p-vector of ones and the absolute value acts elementwise on $\boldsymbol{v}$. The seventh line follows again by Cauchy-Schwarz, and the last again since $\sqrt{x + y} \leq \sqrt{x} + \sqrt{y}$. Now, since $_\epsilon\mu_* < \infty$, and by assumption on $\hat{\boldsymbol{\beta}}^{\mathrm{DY}}, \boldsymbol{\beta}_0$, almost surely,

$$\lim_{N\to\infty} \frac{1}{\sqrt{p}}\|\boldsymbol{a}_{\bullet 1}\|_2 = \lim_{N\to\infty} \frac{1}{\sqrt{p}}\left\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}} - {}_\epsilon\mu_*\boldsymbol{\beta}_0\right\|_2 \leq \lim_{N\to\infty} \frac{1}{\sqrt{p}}\left\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\right\| + {}_\epsilon\mu_*\frac{1}{\sqrt{p}}\|\boldsymbol{\beta}_0\|_2 \quad < C + {}_\epsilon\mu_*\gamma$$

$$\lim_{N\to\infty} \frac{1}{\sqrt{p}}\|\boldsymbol{a}_{\bullet 2}\|_2 = \lim_{N\to\infty} \frac{1}{\sqrt{p}}\|\boldsymbol{\beta}_0\|_2 = \gamma,$$

for some constant $C > 0$, which is independent of t, ϵ. Further, by (S64), almost surely

$$\lim_{N\to\infty} \frac{1}{\sqrt{p}}\|\boldsymbol{b}_{\bullet 1}\|_2 = \lim_{N\to\infty} \frac{1}{\sqrt{p}}\left\|\hat{\beta}^t - {}_\epsilon\mu_*\boldsymbol{\beta}_0\right\|_2 = {}_\epsilon\sigma_*^2.$$

Therefore, (S67) gives that almost surely

$$\lim_{N\to\infty} \frac{1}{p}\sum_{j=1}^{p}\left(1 + \|\boldsymbol{a}_j\|_2 + \|\boldsymbol{b}_j\|_2\right)^2 \leq 1 + (C + {}_\epsilon\mu_*\gamma)\left\{2 + C + {}_\epsilon\mu_*\right\} + {}_\epsilon\sigma_*^2\left\{2 + {}_\epsilon\sigma_*^2\right\} + 4\gamma + 2\left\{C + {}_\epsilon\mu_*\gamma + \gamma\right\}\left\{{}_\epsilon\sigma_*^2 + \gamma\right\}.$$
(S68)

And thus

$$\lim_{\epsilon\to 0}\lim_{t\to\infty}\lim_{N\to\infty} \frac{1}{p}\sum_{j=1}^{p}\left(1 + \|\boldsymbol{a}_j\|_2 + \|\boldsymbol{b}_j\|_2\right)^2 \leq 1 + (C + \mu_*\gamma)\left\{2 + C + \mu_*\right\} + \sigma_*^2\left\{2 + \sigma_*^2\right\} + 4\gamma + 2\left\{C + \mu_*\gamma + \gamma\right\}\left\{\sigma_*^2 + \gamma\right\},$$

where $\lim_{\epsilon \to 0}({}_\epsilon\mu_*, {}_\epsilon\sigma_*) = (\mu_*, \sigma_*)$ follows from Lemma G.6. In conjunction with Theorem B.3 and (S66), this yields that almost surely

$$\lim_{\epsilon \to 0} \lim_{t \to \infty} \lim_{N \to \infty} \frac{1}{p} \sum_{j=1}^{p} \left| \psi\left(\hat{\boldsymbol{\beta}}_j^{\mathrm{DY}} - {}_\epsilon\mu_*\boldsymbol{\beta}_{0,j}, \boldsymbol{\beta}_{0,j}\right) - \psi\left({}_\epsilon\hat{\boldsymbol{\beta}}^t - {}_\epsilon\mu_*\boldsymbol{\beta}_{0,j}, \boldsymbol{\beta}_{0,j}\right) \right| = 0 \,.$$

Similarly,

$$\left| \frac{1}{p} \sum_{j=1}^{p} \psi\left({}_\epsilon\hat{\boldsymbol{\beta}}^t - {}_\epsilon\mu_*\boldsymbol{\beta}_{0,j}, \boldsymbol{\beta}_{0,j}\right) - \psi\left({}_\epsilon\boldsymbol{\nu}^t, \boldsymbol{\beta}_{0,j}\right) \right|$$

$$\leq C \frac{1}{p} \sqrt{\sum_{j=1}^{p} \left(1 + \left\| \begin{bmatrix} {}_\epsilon\hat{\boldsymbol{\beta}}^t - {}_\epsilon\mu_*\boldsymbol{\beta}_{0,j} \\ \boldsymbol{\beta}_{0,j} \end{bmatrix} \right\|_2 + \left\| \begin{bmatrix} {}_\epsilon\boldsymbol{\nu}^t \\ \boldsymbol{\beta}_{0,j} \end{bmatrix} \right\|_2 \right)^2} \left\| {}_\epsilon\hat{\boldsymbol{\beta}}^t - {}_\epsilon\mu_*\boldsymbol{\beta}_0 - {}_\epsilon\boldsymbol{\nu}^t \right\|_2 \,,$$

the right hand side of which is zero almost surely as $N \to \infty$ for all $t, \epsilon > 0$, by Lemma B.4 and (S68). and thus almost surely

$$\lim_{\epsilon \to 0} \lim_{t \to \infty} \lim_{N \to \infty} \left| \frac{1}{p} \sum_{j=1}^{p} \psi\left(\hat{\boldsymbol{\beta}}_j^{\mathrm{DY}} - \mu_*\boldsymbol{\beta}_{0,j}, \boldsymbol{\beta}_{0,j}\right) - \psi\left({}_\epsilon\hat{\boldsymbol{\beta}}_j^{\mathrm{DY}} - {}_\epsilon\mu_*\boldsymbol{\beta}_{0,j}, \boldsymbol{\beta}_{0,j}\right) \right| = 0 \,.$$

Now recall that from Corollary B.2.1, and by the stationarity assumption of the phase transition parameters, $({}_\epsilon\mu_*, {}_\epsilon b_*, {}_\epsilon\sigma_*)$, almost surely

$$\lim_{t \to \infty} \lim_{N \to \infty} \frac{1}{p} \sum_{j=1}^{p} \psi({}_\epsilon\boldsymbol{\nu}_j^t, \boldsymbol{\beta}_{0,j}) = \lim_{N \to \infty} \frac{1}{p} \sum_{j=1}^{p} \psi({}_\epsilon\boldsymbol{\nu}_j^t, \boldsymbol{\beta}_{0,j}) = \mathrm{E}\left[\psi({}_\epsilon\sigma_* G, \bar{\beta})\right] \,,$$

where $G \sim N(0,1) \perp \bar{\beta}$. Since ψ is pseudo-Lipschitz of order two,

$$\left| \mathrm{E}\left[\psi({}_\epsilon\sigma_* G, \bar{\beta})\right] - \mathrm{E}\left[\psi(\sigma_* G, \bar{\beta})\right] \right| \leq \mathrm{E}\left[\left|\psi({}_\epsilon\sigma_* G, \bar{\beta}) - \psi(\sigma_* G, \bar{\beta})\right|\right]$$

$$\leq C \left|{}_\epsilon\sigma_* - \sigma_*\right| \mathrm{E}\left[\left(1 + \left\|[{}_\epsilon\sigma_* G, \bar{\beta}]^\top\right\|_2 + \left\|[\sigma_* G, \bar{\beta}]^\top\right\|_2\right) |G|\right] \,,$$

and thus it follows that

$$\lim_{\epsilon \to 0} \lim_{t \to \infty} \lim_{N \to \infty} \frac{1}{p} \sum_{j=1}^{p} \psi\left({}_\epsilon\boldsymbol{\nu}_j^t, \boldsymbol{\beta}_{0,j}\right) = \mathrm{E}\left[\psi(\sigma_* G, \bar{\beta})\right] \,.$$

In conjunction with (S65), this establishes (S6). Finally, equivalence of stationary solutions (S3)-(S5) with the stationary solutions to (S16)-(S18) is shown in Lemma G.9. This concludes the proof of the Theorem.

C   Proof of Theorem 3.2

Theorem 3.2. *In the setting of Theorem 3.1, and if (κ, γ) are such that the ML estimate exists with probability approaching one, then almost surely*

$$\lim_{n \to \infty} \frac{\left\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}(\alpha)\right\|_2}{\sqrt{n}} = \mathcal{O}(1) \,, \tag{S69}$$

for any $\alpha \in (0,1)$. Furthermore, for any (κ, γ), there exists an $\alpha_0 \in (0,1)$ such that (S69) holds almost surely for all $0 < \alpha < \alpha_0$.

Both parts of Theorem 3.2 are proven in turn.

Theorem C.1. *Under the modelling assumption of Section A, and if $\gamma < g_{MLE}(\kappa)$, where g_{MLE} is the phase transition of Candès and Sur (2020), then almost surely*

$$\lim_{n \to \infty} \left\| \hat{\boldsymbol{\beta}}^{DY} \right\|_2 = \mathcal{O}(\sqrt{n}).$$

Proof. The proof relies on the arguments in the proof of Theorem 4 in the supplementary material of Sur and Candès (2019). All that needs to be done is to show that the DY-prior penalty provides some slackness in the string of inequalities that guarantee boundedness of the estimator.

Write the negative log-likelihood with pseudo-responses $\boldsymbol{y}^*$ as

$$-\log \ell(\boldsymbol{\beta}; \boldsymbol{y}^*, \boldsymbol{X}) = \alpha \sum_{i=1}^{n} \zeta\left(-\tilde{y}_i \boldsymbol{x}_i^\top \boldsymbol{\beta}\right) + (1-\alpha)\left[\sum_{i=1}^{n} -\frac{1}{2}\boldsymbol{x}_i^\top \boldsymbol{\beta} + \zeta(\boldsymbol{x}_i^\top \boldsymbol{\beta})\right],$$

where $\tilde{y}_i = 2y_i - 1$ and $\zeta(x) = \log(1 + e^x)$. Assume that

  i) Given a constant $c > 0$, for $\mathcal{B} = \bigcap_{n > N_0} \{s_{\min}(\boldsymbol{X}) > c\}$ be the event that there exists a N_0 such that for all $n > N_0$, $s_{\min}(\boldsymbol{X}) > c$

  ii) $-\ell(\boldsymbol{\beta}; \boldsymbol{y}^*, \boldsymbol{X}) \leq -\ell(\boldsymbol{0}_p; \boldsymbol{y}^*, \boldsymbol{X}) = n \log(2)$,

  iii) $\frac{\|\boldsymbol{\beta}\|_2}{\sqrt{n}} > \frac{\log(2)}{c\epsilon^2}$

Let $J = \{j \in \{1, \ldots, n\} : \tilde{y}_j x_j^\top \boldsymbol{\beta} < 0\}$. Then note that ii) implies that

$$\sum_{j=1}^{n} \max\left\{-\tilde{y}_j \boldsymbol{x}_j^\top \boldsymbol{\beta}, 0\right\} \leq n \log(2). \tag{S70}$$

Indeed, consider

$$\alpha \sum_{j=1}^{n} \max\left\{-\tilde{y}_j \boldsymbol{x}_j^\top \boldsymbol{\beta}, 0\right\} + (1-\alpha)\left[\sum_{j=1}^{n} -\frac{1}{2}\boldsymbol{x}_j^\top \boldsymbol{\beta} + \zeta(\boldsymbol{x}_j^\top \boldsymbol{\beta})\right] = \alpha \sum_{j \in J} -\tilde{y}_j \boldsymbol{x}_j^\top \boldsymbol{\beta} + (1-\alpha)\left[\sum_{j=1}^{n} -\frac{1}{2}\boldsymbol{x}_j^\top \boldsymbol{\beta} + \zeta(\boldsymbol{x}_j^\top \boldsymbol{\beta})\right]$$

$$\leq \alpha \sum_{j \in J} \zeta\left(-\tilde{y}_j \boldsymbol{x}_j^\top \boldsymbol{\beta}\right) + (1-\alpha)\left[\sum_{j=1}^{n} -\frac{1}{2}\boldsymbol{x}_j^\top \boldsymbol{\beta} + \zeta(\boldsymbol{x}_i^\top \boldsymbol{\beta})\right]$$

$$\leq \alpha \sum_{j=1}^{n} \zeta\left(-\tilde{y}_j \boldsymbol{x}_j^\top \boldsymbol{\beta}\right) + (1-\alpha)\left[\sum_{j=1}^{n} -\frac{1}{2}\boldsymbol{x}_j^\top \boldsymbol{\beta} + \zeta(\boldsymbol{x}_j^\top \boldsymbol{\beta})\right]$$

$$= -\ell(\boldsymbol{\beta}; \boldsymbol{y}^*, \boldsymbol{X}),$$

where it was used that $x < \zeta(x)$ for $x \in \Re$. Now assume that $\boldsymbol{\beta}$ is such that $-\ell(\boldsymbol{\beta}; \boldsymbol{y}^*, \boldsymbol{X}) \leq n \log(2) = -\ell(\boldsymbol{0}_p; \boldsymbol{y}^*, \boldsymbol{X})$. Then for such $\boldsymbol{\beta}$,

$$\alpha \sum_{j \in J} \zeta\left(-\tilde{y}_j \boldsymbol{x}_j^\top \boldsymbol{\beta}\right) \leq n \log(2) - (1-\alpha)\left[\sum_{i=1}^{n} -\frac{1}{2}\boldsymbol{x}_i^\top \boldsymbol{\beta} + \zeta(\boldsymbol{x}_i^\top \boldsymbol{\beta})\right]. \tag{S71}$$

Now note that $(1-\alpha)\left[\sum_{i=1}^{n} -\frac{1}{2}\boldsymbol{x}_i^\top \boldsymbol{\beta} + \zeta(\boldsymbol{x}_i^\top \boldsymbol{\beta})\right]$ is always greater than zero as $\zeta(x) > x$ for $x \in \Re$

31

and that this expression is minimised for $\boldsymbol{\beta} = \boldsymbol{0}_p$. Hence, (S71) can be upper bounded as

$$\alpha \sum_{j \in J} \zeta \left(-\tilde{y}_j \boldsymbol{x}_j^\top \boldsymbol{\beta} \right) \leq n \log(2) - (1 - \alpha) \left[\sum_{i=1}^n -\frac{1}{2} \boldsymbol{x}_i^\top \boldsymbol{\beta} + \zeta(\boldsymbol{x}_i^\top \boldsymbol{\beta}) \right]$$

$$\leq n \log(2) - (1 - \alpha) \left[\sum_{i=1}^n -\frac{1}{2} \boldsymbol{x}_i^\top \boldsymbol{0}_p + \zeta(\boldsymbol{0}_p \boldsymbol{\beta}) \right]$$

$$= n \log(2) - (1 - \alpha) n \log(2)$$

$$= \alpha n \log(2) \,,$$

which establishes (S70) since then by definition of J and as $\zeta(x) > x$,

$$\sum_{j=1}^n \max \left\{ -\tilde{y}_j \boldsymbol{x}_j^\top \boldsymbol{\beta}, 0 \right\} = \sum_{j \in J} -\tilde{y}_j \boldsymbol{x}_j^\top \boldsymbol{\beta} \leq \sum_{j \in J} \zeta \left(-\tilde{y}_j \boldsymbol{x}_j^\top \boldsymbol{\beta} \right) \leq n \log(2) \,.$$

Next, note that conditional on event $\mathcal{B}$ from assumptions i) & iii) it follows that

$$n \log(2) \leq \frac{n \log(2)}{c} s_{\min}(\boldsymbol{X}) \leq \frac{n \log(2)}{c} \frac{\|\boldsymbol{X}\boldsymbol{\beta}\|_2}{\|\boldsymbol{\beta}\|_2} = \frac{n \log(2)}{c} \frac{\|\tilde{\boldsymbol{y}} \circ \boldsymbol{X}\boldsymbol{\beta}\|_2}{\|\boldsymbol{\beta}\|_2} \leq \epsilon^2 \sqrt{n} \|\tilde{\boldsymbol{y}} \circ \boldsymbol{X}\boldsymbol{\beta}\|_2 \,,$$

where the first inequality holds by event $\mathcal{B}$, the second by the Courant-Fischer characterisation of eigenvalues (e.g. Magnus and Neudecker 2019, Section 1.8, Theorem 7), the next equality since $\tilde{\boldsymbol{y}} \in \{-1, 1\}$ and the definition of the Hadamard product $\circ$ and the ℓ_2-norm, and the last inequality follows by iii).

Note that by definition of $\hat{\boldsymbol{\beta}}^{\mathrm{DY}}$, it must hold that $-\ell(\hat{\boldsymbol{\beta}}^{\mathrm{DY}}; \boldsymbol{y}^*, \boldsymbol{X}) \leq n \log(2)$. Hence, conditional on event $\mathcal{B}$, $\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\|_2 / \sqrt{n} > \log(2)/(c\epsilon^2)$ implies that $\tilde{\boldsymbol{y}} \circ \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\mathrm{DY}}$ must fall in the cone

$$\mathcal{A} = \left\{ \boldsymbol{u} \in \Re^n : \sum_{j=1}^n \max \left\{ -u_j, 0 \right\} \leq \epsilon^2 \sqrt{n} \|\boldsymbol{u}\|_2 \right\} \,.$$

Therefore

$$\Pr \left(\frac{\left\| \hat{\boldsymbol{\beta}}^{\mathrm{DY}} \right\|_2}{\sqrt{n}} > \frac{\log(2)}{c\epsilon^2} \middle| \mathcal{B} \right) \leq \Pr \left(\tilde{\boldsymbol{y}} \circ \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\mathrm{DY}} \in \mathcal{A} \middle| \mathcal{B} \right) \,.$$

Next, note that by a straightforward application of Theorem 2.6 in Rudelson and Vershynin (2010), it follows that for any κ, there exist constants $c_1, c_2 > 0$ such that

$$\Pr \left(s_{\min}(\boldsymbol{X}) > c_1 \right) \geq 1 - 2 \exp \left\{ -c_2 n \right\} \,.$$

Therefore, letting $\mathcal{C}_n$ be the event that $s_{\min}(\boldsymbol{X}) < c_1$ for $\boldsymbol{X} \in \Re^{n \times p}$, the first Borel-Cantelli Lemma gives that the event $\Pr(\mathcal{C}_n \text{ infinitely often}) = 0$. Hence, there must exist a N_0 such that $\Pr \left(\bigcap_{n > N_0} s_{\min}(\boldsymbol{X}) > c_1 \right) = 1$. Thus, letting $\mathcal{B} = \bigcap_{n > N_0} s_{\min}(\boldsymbol{X}) > c_1$, it follows that for $n > N_0$,

$\epsilon > 0$,

$$\Pr\left(\frac{\left\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\right\|_2}{\sqrt{n}} > \frac{\log(2)c_1}{\epsilon^2}\right) = \Pr\left(\frac{\left\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\right\|_2}{\sqrt{n}} > \frac{\log(2)c_1}{\epsilon^2}\middle|\mathcal{B}\right)\Pr(\mathcal{B})$$

$$+ \Pr\left(\frac{\left\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\right\|_2}{\sqrt{n}} > \frac{\log(2)c_1}{\epsilon^2}\middle|\neg\mathcal{B}\right)(1 - \Pr(\mathcal{B}))$$

$$= \Pr\left(\frac{\left\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\right\|_2}{\sqrt{n}} > \frac{\log(2)c_1}{\epsilon^2}\middle|\mathcal{B}\right) \cdot 1 + 0$$

$$= \Pr\left(\frac{\left\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\right\|_2}{\sqrt{n}} > \frac{\log(2)c_1}{\epsilon^2}\middle|\mathcal{B}\right)$$

A similar argument shows that $\Pr\left(\tilde{\boldsymbol{y}} \circ \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\mathrm{DY}} \in \mathcal{A}|\mathcal{B}\right) = \Pr\left(\tilde{\boldsymbol{y}} \circ \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\mathrm{DY}} \in \mathcal{A}\right)$. Thus, for $n > N_0$,

$$\Pr\left(\frac{\left\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\right\|_2}{\sqrt{n}} > \frac{\log(2)c_1}{\epsilon^2}\right) \leq \Pr\left(\tilde{\boldsymbol{y}} \circ \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\mathrm{DY}} \in \mathcal{A}\right)$$

$$\leq \Pr\left(\{\tilde{\boldsymbol{y}} \circ \boldsymbol{X}\boldsymbol{b} : \boldsymbol{b} \in \Re^p\} \cap \mathcal{A} \neq \{\boldsymbol{0}_n\}\right)$$

This event is shown to have probability at most $Cn^{-\delta}$ for some constants $C > 0, \delta > 1$ in the proof of Theorem 4 in the supplementary material of Sur and Candès (2019) (see also Sur *et al.* 2019, Theorem 7 for more details).

The almost sure part of the claim follows again by the first Borel-Cantelli Lemma. This concludes the proof. □

Theorem C.2. *In the setting of Section A, for any pair (κ, γ), there exists an $\alpha_0 \in (0,1)$ such that for all $0 < \alpha < \alpha_0$,*

$$\left\|\hat{\boldsymbol{\beta}}^{DY}(\alpha)\right\|_2 = \mathcal{O}(\sqrt{n})$$

almost surely for the corresponding MDYPL estimator, $\hat{\boldsymbol{\beta}}^{DY}(\alpha)$, with shrinkage parameter α.

Proof. It shall be argued that for any κ, γ, there exists a radius $r_0 = c\sqrt{n}$ for some constant $c > 0$ and a cutoff $\alpha_0 \in (0,1)$ such that for all $\alpha < \alpha_0$, the maximiser of $\ell(\boldsymbol{\beta}; \boldsymbol{y}^*, \boldsymbol{X})$ as defined in (S2) lies in the set $S = \{\boldsymbol{\beta} \in \Re^p : \|\boldsymbol{\beta}\|_2 \leq r_0\}$ eventually almost surely. Towards this end, let $p(\boldsymbol{\beta}) = \frac{1-\alpha}{\alpha}\sum_{i=1}^n \frac{1}{2}\boldsymbol{x}_i^\top\boldsymbol{\beta} - \zeta(\boldsymbol{x}_i^\top\boldsymbol{\beta})$ be the log of the prior (S1) with $\boldsymbol{\beta}_P = \boldsymbol{0}_p$ bar the normalizing constant and fix any $\kappa \in (0,1)$, $\gamma > 0$. By Lemma C.1, there exist $c > 0$, $\alpha_0 \in (0,1)$ such that for all $0 < \alpha < \alpha_0$, such that

$$\sup_{\substack{\boldsymbol{\beta}\in\Re^p: \\ \|\boldsymbol{\beta}\|_2 > c\sqrt{n}}} p(\boldsymbol{\beta}) < -\frac{1}{\alpha}n\log(2), \tag{S72}$$

with probability at least $1 - c_1\exp\{-c_2 n\}$ for some universal constants c_1, c_2 and $n > n_0$. Since the log-likelihood $\ell(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{X})$ from a logistic regression model is always bounded from above by zero, it follows that, upon the event of (S72), $\ell(\boldsymbol{\beta}; \boldsymbol{y}^*, \boldsymbol{X}) = \ell(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{X}) + p(\boldsymbol{\beta}) < -n\log(2)/\alpha = \ell(\boldsymbol{0}_p; \boldsymbol{y}^*, \boldsymbol{X})$. On the other hand, $\ell(\boldsymbol{\beta}; \boldsymbol{y}^*, \boldsymbol{X})$ obtains a maximum over S since it is continuous and S is compact (e.g. Rudin (1976, Theorem 4.16)). Further, since for any $\boldsymbol{\beta} \notin S$, $\ell(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{X}) < \ell(\boldsymbol{0}_p; \boldsymbol{y}, \boldsymbol{X})$ and $\boldsymbol{0}_p \in S$, the maximiser over S is the global maximiser of $\ell(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{X})$, which thus must be bounded by $c\sqrt{n}$ with probability at least $1 - c_1\exp\{-c_2 n\}$. The almost sure part of the statement follows by

33

the Borel-Cantelli Lemma. This concludes the proof. □

Lemma C.1. *Let $p(\boldsymbol{\beta}) = \frac{1-\alpha}{\alpha} \sum_{i=1}^{n} \frac{1}{2} \boldsymbol{x}_i^\top \boldsymbol{\beta} - \zeta(\boldsymbol{x}_i^\top \boldsymbol{\beta})$ be the log of the prior (S1) with $\boldsymbol{\beta}_P = \boldsymbol{0}_p$ bar the normalizing constant. Then for any $\kappa \in (0,1)$, $\gamma > 0$, there exist $c > 0$, $\alpha_0 \in (0,1)$, n_0, such that for all $0 < \alpha < \alpha$, $n > n_0$*

$$\sup_{\substack{\boldsymbol{\beta} \in \Re^p : \\ \|\boldsymbol{\beta}\|_2 > c\sqrt{n}}} p(\boldsymbol{\beta}) < -\frac{1}{\alpha} n \log(2),$$

with probability at least $1 - c_1 \exp\{-c_2 n\}$ for some universal constants c_1, c_2.

Proof. First write $p(\boldsymbol{\beta}) = p(r\boldsymbol{u})$ for $r = \|\boldsymbol{\beta}\|_2, \boldsymbol{u} \in S^{p-1}$ and note that $p(r\boldsymbol{u})$ is monotonically decreasing in r. Indeed,

$$\frac{\partial}{\partial r} p(r\boldsymbol{u}) = \sum_{i=1}^{n} \boldsymbol{x}_i^\top \boldsymbol{u} \left(\frac{1}{2} - \zeta'(r\boldsymbol{x}_i^\top \boldsymbol{u}) \right),$$

and the statement is trivial for $r = 0$. Hence assume $r > 0$. If $\boldsymbol{x}_i^\top \boldsymbol{u} > 0$, then by elementary properties of $\zeta(\cdot)$, $\zeta'(r\boldsymbol{x}_i^\top \boldsymbol{u}) > 1/2$. On the other hand if $\boldsymbol{x}_i^\top \boldsymbol{u} < 0$, $\zeta'(r\boldsymbol{x}_i^\top \boldsymbol{u}) < 1/2$ and if $\boldsymbol{x}_i^\top \boldsymbol{u} = 0$, $\zeta'(r\boldsymbol{x}_i^\top \boldsymbol{u}) = 1/2$. Thus, for $r > 0$, $\frac{\partial}{\partial r} p(r\boldsymbol{u}) < 0$ except on a set of measure zero, on which it is ≤ 0. Hence it is sufficient to establish that there exists a constant $c > 0$, such that for $r = c\sqrt{n}$

$$\sup_{\boldsymbol{u} \in S^{p-1}} p(r\boldsymbol{u}) < -\frac{1}{\alpha} n \log(2), \tag{S73}$$

with sufficiently high probability. A second order Taylor expansion of $\zeta(r\boldsymbol{x}_i^\top \boldsymbol{u})$ around 0 gives

$$\zeta(r\boldsymbol{x}_i^\top \boldsymbol{u}) = \zeta(0) + \zeta'(0)r\boldsymbol{x}_i^\top \boldsymbol{u} + \frac{1}{2}\zeta''(\gamma_i)(r\boldsymbol{x}_i^\top \boldsymbol{u})^2$$

$$= \log(2) + \frac{1}{2}r\boldsymbol{x}_i^\top \boldsymbol{u} + \frac{1}{2}\zeta''(\gamma_i)(r\boldsymbol{x}_i^\top \boldsymbol{u})^2,$$

where γ_i lies between 0 and $r\boldsymbol{x}_i^\top \boldsymbol{u}$. Hence,

$$p(r\boldsymbol{u}) = \frac{1-\alpha}{\alpha} \sum_{i=1}^{n} \frac{1}{2} r\boldsymbol{x}_i^\top \boldsymbol{u} - \zeta(r\boldsymbol{x}_i^\top \boldsymbol{u})$$

$$= \frac{1-\alpha}{\alpha} \sum_{i=1}^{n} \left\{ \frac{1}{2} r\boldsymbol{x}_i^\top \boldsymbol{u} - \log(2) - \frac{1}{2} r\boldsymbol{x}_i^\top \boldsymbol{u} - \frac{1}{2}\zeta''(\gamma_i)(r\boldsymbol{x}_i^\top \boldsymbol{u})^2 \right\}$$

$$= -\frac{1-\alpha}{\alpha} n \log(2) - \frac{1-\alpha}{2\alpha} \sum_{i=1}^{n} \zeta''(\gamma_i)(r\boldsymbol{x}_i^\top \boldsymbol{u})^2$$

$$\leq -\frac{1-\alpha}{\alpha} n \log(2) - \frac{1-\alpha}{2\alpha} \sum_{i=1}^{n} \zeta''(r\boldsymbol{x}_i^\top \boldsymbol{u})(r\boldsymbol{x}_i^\top \boldsymbol{u})^2$$

$$= -\frac{1-\alpha}{\alpha} n \log(2) - \frac{1-\alpha}{2\alpha} (r\boldsymbol{u})^\top \boldsymbol{X}^\top \boldsymbol{W}\{r\boldsymbol{u}\} \boldsymbol{X}(r\boldsymbol{u}),$$

where $\boldsymbol{W}\{r\boldsymbol{u}\}$ is a diagonal matrix with $[\boldsymbol{W}\{r\boldsymbol{u}\}]_{ii} = \zeta''(r\boldsymbol{x}_i^\top \boldsymbol{u})$. The inequality follows since $\zeta''(\cdot)$ is positive, symmetric around zero $\zeta''(x) = \zeta''(-x)$, and strictly decreasing away from zero so that $|x| > |y|$ implies $\zeta''(y) > \zeta''(x)$. Since γ_i lies between 0 and $r\boldsymbol{x}_i^\top \boldsymbol{u}$, $|\gamma_i| \leq |r\boldsymbol{x}_i^\top \boldsymbol{u}|$ and

34

$\zeta''(\gamma_i) > \zeta''(r\boldsymbol{x}_i^\top \boldsymbol{u})$ as required. Hence,

$$
\begin{aligned}
\sup_{\boldsymbol{u} \in S^{p-1}} p(r\boldsymbol{u}) &\le -\frac{1-\alpha}{\alpha} n \log(2) + \frac{1-\alpha}{2\alpha} \sup_{\boldsymbol{u} \in S^{p-1}} - (r\boldsymbol{u})^\top \boldsymbol{X}^\top \boldsymbol{W}\{r\boldsymbol{u}\} \boldsymbol{X}(r\boldsymbol{u}) \\
&= -\frac{1-\alpha}{\alpha} n \log(2) - \frac{1-\alpha}{2\alpha} \inf_{\boldsymbol{u} \in S^{p-1}} (r\boldsymbol{u})^\top \boldsymbol{X}^\top \boldsymbol{W}\{r\boldsymbol{u}\} \boldsymbol{X}(r\boldsymbol{u}) \\
&\le -\frac{1-\alpha}{\alpha} n \log(2) - \frac{1-\alpha}{2\alpha} \inf_{\boldsymbol{v} \in S^{p-1}} \inf_{\boldsymbol{u} \in S^{p-1}} (r\boldsymbol{u})^\top \boldsymbol{X}^\top \boldsymbol{W}\{r\boldsymbol{v}\} \boldsymbol{X}(r\boldsymbol{u}) \\
&= -\frac{1-\alpha}{\alpha} n \log(2) - \frac{1-\alpha}{2\alpha} r^2 \inf_{\boldsymbol{v} \in S^{p-1}} \lambda_{\min}\left(\boldsymbol{X}^\top \boldsymbol{W}\{r\boldsymbol{v}\} \boldsymbol{X}\right),
\end{aligned}
$$

and thus, to establish (S73) it is sufficient that

$$
\frac{1-\alpha}{2\alpha} r^2 \inf_{\boldsymbol{v} \in S^{p-1}} \lambda_{\min}\left(\boldsymbol{X}^\top \boldsymbol{W}\{r\boldsymbol{v}\} \boldsymbol{X}\right) > n \log(2).
$$

By Lemma 7 of Sur *et al.* (2019), there exists a $\epsilon_0 > 0$ such that for all $0 < \epsilon \le \epsilon_0$,

$$
\inf_{\boldsymbol{v} \in S^{p-1}} \lambda_{\min}\left(\boldsymbol{X}^\top \boldsymbol{W}\{r\boldsymbol{v}\} \boldsymbol{X}\right) \ge \zeta''\left(\frac{3r}{\sqrt{\epsilon n}}\right)\left(\sqrt{1-\epsilon} - \sqrt{\frac{p}{n}} - 2\sqrt{\frac{H(\epsilon)}{1-\epsilon}}\right)^2,
$$

with probability at least $1 - 2\exp\{-nH(\epsilon)\} - 2\exp\{-n/\epsilon\}$, where $H(\epsilon) = -\epsilon \log(\epsilon) - (1-\epsilon)\log(1-\epsilon)$. Hence, since $r = c\sqrt{n}$, for $0 < \epsilon \le \epsilon_0$, we can find constants $c_1(\epsilon, \kappa), c_2(\epsilon, \kappa), C(\epsilon, \kappa) > 0$, and $n_0(\kappa, \epsilon)$ such that for all $n > n_0(\kappa, \epsilon)$,

$$
\inf_{\boldsymbol{v} \in S^{p-1}} \lambda_{\min}\left(\boldsymbol{X}^\top \boldsymbol{W}\{r\boldsymbol{v}\} \boldsymbol{X}\right) \ge \zeta''\left(\frac{3c}{\sqrt{\epsilon}}\right) C(\epsilon, \kappa),
$$

with probability at least $1 - c_1(\epsilon, \kappa)\exp\{-c_2(\epsilon, \kappa)n\}$. On that event, for α small enough

$$
\frac{1-\alpha}{2\alpha} r^2 \inf_{\boldsymbol{v} \in S^{p-1}} \lambda_{\min}\left(\boldsymbol{X}^\top \boldsymbol{W}\{r\boldsymbol{v}\} \boldsymbol{X}\right) \ge \frac{1-\alpha}{2\alpha} c^2 n \zeta''\left(\frac{3c}{\sqrt{\epsilon}}\right) C(\kappa, \epsilon) > n \log(2).
$$

This concludes the proof. □

Remark 1. Note that the proof of Lemma C.1 does not depend on the specification of γ. Hence, the threshold value α_0, which guarantees boundedness of $\hat{\boldsymbol{\beta}}^{\mathrm{DY}}$ in ℓ_2-norm also depends solely on κ. A more refined analysis might be able to produce a concrete upper bound of the cutoff value $\alpha_0(\kappa)$.

D   Proof of Theorem 3.3

Recall that the setup of Theorems 3.3-3.5 is the moderately high-dimensional setting with i.i.d. Gaussian covariates with arbitrary covariance structure. In particular, $\boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{0}_p, \boldsymbol{\Sigma})$ for some positive definite matrix $\boldsymbol{\Sigma}$ and signal $\boldsymbol{\beta}_0 \in \Re^p$ such that $\lim_{n \to \infty} \mathrm{var}(\boldsymbol{x}_i^\top \boldsymbol{\beta}_0) = \boldsymbol{\beta}_0^\top \boldsymbol{\Sigma} \boldsymbol{\beta}_0 = \gamma^2$ for $\gamma \in \Re_{>0}^p$ and where $\lim_{n \to \infty} p/n \to \kappa \in (0, 1)$.

Theorem 3.3. *Assume that (α, κ, γ) are such that the conditions of Theorem 3.1 are met. Let $\hat{\boldsymbol{\beta}}^{\mathrm{DY}}$ be the MDYPL estimator for a logistic regression model with i.i.d. $\mathcal{N}(\boldsymbol{0}_p, \boldsymbol{\Sigma})$ covariates, with $\limsup_{n \to \infty} \lambda_{\max}(\boldsymbol{\Sigma})/\lambda_{\min}(\boldsymbol{\Sigma}) < \infty$, and signal $\boldsymbol{\beta}_0$. Then for any $t \in \Re$,*

$$\frac{1}{p}\sum_{j=1}^{p}\mathbb{1}\left\{\sqrt{n}\tau_j\frac{\hat{\beta}_j^{DY}-\mu_*\beta_{0,j}}{\sigma_*}\leq t\right\}\xrightarrow{p}\Phi(t),$$

where $\Phi(\cdot)$ is the CDF of a standard normal random variable, $\mathbb{1}(\cdot)$ is the indicator function and $\tau_j^2 = var(\boldsymbol{x}_{ij}|\boldsymbol{x}_{i-j})$ is the conditional variance of the jth predictor given all others. Further, if

$$\frac{1}{p}\sum_{j=1}^{p}\delta_{\sqrt{n}\tau_j\boldsymbol{\beta}_{0,j}}\xrightarrow{d}\pi_{\bar{\beta}}\quad and \quad \frac{1}{p}\sum_{j=1}^{p}n\tau_j^2\beta_{0,j}^2\xrightarrow{p}E(\bar{\beta}^2),$$

for some distribution $\pi_{\bar{\beta}}$ with finite second moment and $\bar{\beta}\sim\pi_{\bar{\beta}}$, where $\delta_{(\cdot)}$ is the Dirac delta function, then for any pseudo-Lipschitz function $\psi:\Re^2\to\Re$ of order 2,

$$\frac{1}{p}\sum_{j=1}^{p}\psi\left(\sqrt{n}\tau_j\left(\hat{\beta}_j^{DY}-\mu_*\boldsymbol{\beta}_{0,j}\right),\sqrt{n}\tau_j\boldsymbol{\beta}_{0,j}\right)\xrightarrow{p}E\left[\psi(\sigma_*G,\bar{\beta})\right],\quad as\ n\to\infty,$$

for $G\sim N(0,1)$ independent of $\bar{\beta}$.

The proof of Theorem 3.3 is a replication of the arguments in the proof of Theorem 3.2 in Zhao *et al.* (2022) with the substitutions that are outlined in F. No claim of originality is made.

The following Lemma, which is a generalization of Lemma E.2 to arbitrary covariance structures, is the starting point of the analysis. Let $\langle\boldsymbol{u},\boldsymbol{v}\rangle_{\boldsymbol{\Sigma}}=\boldsymbol{u}^\top\boldsymbol{\Sigma}\boldsymbol{v}$, and $\|\boldsymbol{u}\|_{\boldsymbol{\Sigma}}^2=\langle\boldsymbol{u},\boldsymbol{u}\rangle_{\boldsymbol{\Sigma}}$. Then write $\gamma_n^2=\|\boldsymbol{\beta}_0\|_{\boldsymbol{\Sigma}}^2=var(\boldsymbol{x}_i^\top\boldsymbol{\beta}_0)$ for $\boldsymbol{x}_i\sim\mathcal{N}(\boldsymbol{0}_p,\boldsymbol{\Sigma})$.

Lemma D.1. *Let $\hat{\boldsymbol{\beta}}^{DY}$ be the MDYPL estimator from $\mathcal{N}(\boldsymbol{0}_p,\boldsymbol{\Sigma})$ covariates and signal $\boldsymbol{\beta}_0$. Define*

$$\boldsymbol{\xi}=\boldsymbol{Z}-\left\langle\boldsymbol{Z},\frac{\boldsymbol{\beta}_0}{\gamma_n}\right\rangle_{\boldsymbol{\Sigma}}\frac{\boldsymbol{\beta}_0}{\gamma_n},$$

for $\boldsymbol{Z}\sim\mathcal{N}(\boldsymbol{0}_p,\boldsymbol{\Sigma}^{-1})$ independent of everything else. Further, let

$$\tilde{\mu}_n=\frac{\langle\hat{\boldsymbol{\beta}}^{DY},\boldsymbol{\beta}_0\rangle_{\boldsymbol{\Sigma}}}{\|\boldsymbol{\beta}_0\|_{\boldsymbol{\Sigma}}^2},\quad\tilde{\sigma}_n^2=\left\|\hat{\boldsymbol{\beta}}^{DY}\right\|_{\boldsymbol{\Sigma}}^2-\frac{\langle\hat{\boldsymbol{\beta}}^{DY},\boldsymbol{\beta}_0\rangle_{\boldsymbol{\Sigma}}^2}{\|\boldsymbol{\beta}_0\|_{\boldsymbol{\Sigma}}^2}.$$

Then

$$\frac{\hat{\boldsymbol{\beta}}^{DY}-\tilde{\mu}_n\boldsymbol{\beta}_0}{\tilde{\sigma}_n}\stackrel{d}{=}\frac{\boldsymbol{\xi}}{\|\boldsymbol{\xi}\|_{\boldsymbol{\Sigma}}}.$$

Proof. The proof is similar to the proof of Proposition B.1 in the proof of Theorem 3.2 of Zhao *et al.* (2022) and thus omitted. □

Lemma D.1 immediately yields the equivalence (S74) below. Define $\boldsymbol{T}^{\text{approx}}$ as the p-vector with elements

$$\boldsymbol{T}_j^{\text{approx}}=\tau_j\sqrt{n}\frac{\hat{\boldsymbol{\beta}}_j^{\text{DY}}-\tilde{\mu}_n\boldsymbol{\beta}_{0,j}}{\tilde{\sigma}_n/\sqrt{\kappa}},$$

which ought to approximate

$$\boldsymbol{T}_j=\tau_j\sqrt{n}\frac{\hat{\boldsymbol{\beta}}_j^{\text{DY}}-\mu_*\boldsymbol{\beta}_{0,j}}{\sigma_*}.$$

Further, let $\boldsymbol{Z}_j^{\text{scaled}}=\tau_j\boldsymbol{Z}_j$, so that

$$\boldsymbol{Z}^{\text{scaled}}\sim\mathcal{N}(\boldsymbol{0}_p,\boldsymbol{R}),\quad\boldsymbol{R}_{ij}=\tau_i\tau_j[\boldsymbol{\Sigma}^{-1}]_{ij},$$

and finally let $\tilde{\boldsymbol{Z}}^{\text{scaled}}$ be the p-vector with entries

$$\tilde{\boldsymbol{Z}}_j^{\text{scaled}} = \left(\boldsymbol{Z}_j^{\text{scaled}} - \left\langle \boldsymbol{Z}, \frac{\boldsymbol{\beta}_0}{\gamma_n} \right\rangle_{\boldsymbol{\Sigma}} \frac{\tau_j \boldsymbol{\beta}_{0,j}}{\gamma_n} \right) \frac{\sqrt{p}}{\|\boldsymbol{\xi}\|_{\boldsymbol{\Sigma}}}.$$

Then

$$\boldsymbol{T}^{\text{approx}} \stackrel{d}{=} \tilde{\boldsymbol{Z}}^{\text{scaled}}. \tag{S74}$$

The remainder of the proof is concerned with showing that $\boldsymbol{T} \approx \boldsymbol{T}^{\text{approx}}$, $\tilde{\boldsymbol{Z}}^{\text{scaled}} \approx \boldsymbol{Z}^{\text{scaled}}$ in an adequate sense and that the empirical CDF of $\boldsymbol{Z}^{\text{scaled}}$ converges to that of a standard normal.

> **Lemma D.2.** *If* $\limsup\limits_{n \to \infty} \lambda_{\max}(\boldsymbol{\Sigma})/\lambda_{\min}(\boldsymbol{\Sigma}) < \infty$, *then for any univariate Lipschitz function* ψ,
>
> $$\frac{1}{p} \sum_{j=1}^{p} \left\{ \psi(\tilde{\boldsymbol{Z}}_j^{\text{scaled}}) - \psi(\boldsymbol{Z}_j^{\text{scaled}}) \right\} \stackrel{p}{\longrightarrow} 0, \quad \frac{1}{p} \sum_{j=1}^{p} \left\{ \psi(\boldsymbol{T}_j^{\text{approx}}) - \psi(\boldsymbol{T}_j) \right\} \stackrel{p}{\longrightarrow} 0.$$

Proof. First consider the pair $\tilde{\boldsymbol{Z}}^{\text{scaled}}$, $\boldsymbol{Z}^{\text{scaled}}$ and let L be the Lipschitz constant of ψ. By the triangle inequality and the definition of $\tilde{\boldsymbol{Z}}^{\text{scaled}}$, $\boldsymbol{Z}^{\text{scaled}}$,

$$\left| \frac{1}{p} \sum_{j=1}^{p} \left\{ \psi(\tilde{\boldsymbol{Z}}_j^{\text{scaled}}) - \psi(\boldsymbol{Z}_j^{\text{scaled}}) \right\} \right| \leq \frac{1}{p} \sum_{j=1}^{p} \left| \psi(\tilde{\boldsymbol{Z}}_j^{\text{scaled}}) - \psi(\boldsymbol{Z}_j^{\text{scaled}}) \right|$$

$$\leq L \frac{1}{p} \sum_{j=1}^{p} \left| \boldsymbol{Z}_j^{\text{scaled}} \left(1 - \frac{\sqrt{p}}{\|\boldsymbol{\xi}\|_{\boldsymbol{\Sigma}}} \right) + \left\langle \boldsymbol{Z}, \frac{\boldsymbol{\beta}_0}{\gamma_n} \right\rangle_{\boldsymbol{\Sigma}} \frac{\tau_j \boldsymbol{\beta}_{0,j}}{\gamma_n} \frac{\sqrt{p}}{\|\boldsymbol{\xi}\|_{\boldsymbol{\Sigma}}} \right|$$

$$\leq L \left| 1 - \frac{\sqrt{p}}{\|\boldsymbol{\xi}\|_{\boldsymbol{\Sigma}}} \right| \frac{1}{p} \sum_{j=1}^{p} \left| \boldsymbol{Z}_j^{\text{scaled}} \right| + \left| \frac{1}{\gamma_n} \left\langle \boldsymbol{Z}, \frac{\boldsymbol{\beta}_0}{\gamma_n} \right\rangle_{\boldsymbol{\Sigma}} \right| \frac{L\sqrt{p}}{\|\boldsymbol{\xi}\|_{\boldsymbol{\Sigma}}} \frac{1}{p} \sum_{j=1}^{p} |\tau_j \boldsymbol{\beta}_{0,j}|. \tag{S75}$$

Arguments similar to Lemma G.10 establish that $\sqrt{p}/\|\boldsymbol{\xi}\|_{\boldsymbol{\Sigma}} \to 1$ almost surely as $n \to \infty$. Further note that $\langle \boldsymbol{Z}, \boldsymbol{\beta}_0/\gamma_n \rangle_{\boldsymbol{\Sigma}}/\gamma_n$ is a univariate Gaussian with $\mathcal{O}(1)$ variance so that it is itself $\mathcal{O}_p(1)$. Hence, it is sufficient to show that

$$\frac{1}{p} \sum_{j=1}^{p} \left| \boldsymbol{Z}_j^{\text{scaled}} \right| = \mathcal{O}_p(1), \quad \frac{1}{p} \sum_{j=1}^{p} |\tau_j \boldsymbol{\beta}_{0,j}| = o(1). \tag{S76}$$

For the first note that $\mathrm{E}[|\boldsymbol{Z}_j^{\text{scaled}}|]^2 \leq \mathrm{E}[\boldsymbol{Z}_j^{\text{scaled}2}] = \tau_j^2 [\boldsymbol{\Sigma}^{-1}]_{jj} \mathrm{E}[\boldsymbol{Z}_j^2] = 1$ by definition of τ_j. Now Markov's inequality (e.g. Vershynin (2018, Proposition 1.2.4)) yields that for $\epsilon > 0$,

$$\Pr\left(\frac{1}{p} \sum_{j=1}^{p} \left| \boldsymbol{Z}_j^{\text{scaled}} \right| \geq \frac{1}{\epsilon} \right) \leq \epsilon \mathrm{E}\left[\frac{1}{p} \sum_{j=1}^{p} \left| \boldsymbol{Z}_j^{\text{scaled}} \right| \right] \leq \epsilon, \tag{S77}$$

so that $\frac{1}{p} \sum_{j=1}^{p} \left| \boldsymbol{Z}_j^{\text{scaled}} \right| = \mathcal{O}_p(1)$ indeed. Now, by the Cauchy-Schwarz inequality,

$$\frac{1}{p} \sum_{j=1}^{p} |\tau_j \boldsymbol{\beta}_{0,j}| = \frac{1}{p} \sum_{j=1}^{p} |\boldsymbol{\theta}_{0,j}| \leq \frac{1}{\sqrt{p}} \|\boldsymbol{\theta}_0\|_2 = \mathcal{O}(p^{-1/2}), \tag{S78}$$

where $\boldsymbol{\theta}_0 = \boldsymbol{L}^\top \boldsymbol{\beta}_0$ for $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top$, and by assumption $\|\boldsymbol{\theta}_0\|_2^2 = \mathrm{var}(\boldsymbol{x}_i^\top \boldsymbol{\beta}_0) \to \gamma^2$ as $n \to \infty$, for $\boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{0}_p, \boldsymbol{\Sigma})$. Hence, (S75) is indeed $o(1)$. Next consider $\boldsymbol{T}, \boldsymbol{T}^{\text{approx}}$.

$$\left| \frac{1}{p} \sum_{j=1}^{p} \psi(\boldsymbol{T}_j^{\text{approx}}) - \psi(\boldsymbol{T}_j) \right| \le \frac{1}{p} \sum_{j=1}^{p} \left| \psi(\boldsymbol{T}_j^{\text{approx}}) - \psi(\boldsymbol{T}_j) \right|$$

$$\le \frac{L}{\sigma_*} \sqrt{\frac{n}{p}} \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \left| \tau_j \hat{\boldsymbol{\beta}}_j^{\text{DY}} \left(1 - \frac{\sqrt{\kappa}\sigma_*}{\tilde{\sigma}_n} \right) + \tau_j \boldsymbol{\beta}_{0,j} \left(\frac{\sqrt{\kappa}\sigma_*}{\tilde{\sigma}_n} \tilde{\mu}_n - \mu_* \right) \right|$$

$$\le \frac{L}{\sigma_*} \sqrt{\frac{n}{p}} \left\{ \left| 1 - \frac{\sqrt{\kappa}\sigma_*}{\tilde{\sigma}_n} \right| \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \left| \tau_j \hat{\boldsymbol{\beta}}_j^{\text{DY}} \right| + \left| \frac{\sqrt{\kappa}\sigma_*}{\tilde{\sigma}_n} \tilde{\mu}_n - \mu_* \right| \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \left| \tau_j \boldsymbol{\beta}_{0,j} \right| \right\} \, .$$
(S79)

Now, it follows readily from Lemma E.1 and Lemma E.3 that $\tilde{\mu}_n \to \mu_*$ and $\tilde{\sigma}_n \to \sigma_*$ almost surely as $n \to \infty$. Indeed, note that by Lemma E.1, one can write

$$\tilde{\mu}_n = \frac{(\hat{\boldsymbol{\beta}}^{\text{DY}})^\top \boldsymbol{\Sigma} \boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^\top \boldsymbol{\Sigma} \boldsymbol{\beta}_0} = \frac{(\boldsymbol{L}^\top \hat{\boldsymbol{\beta}}^{\text{DY}})^\top (\boldsymbol{L}^\top \boldsymbol{\beta}_0)}{(\boldsymbol{L}^\top \boldsymbol{\beta}_0)^\top (\boldsymbol{L}^\top \boldsymbol{\beta}_0)} = \frac{\langle \hat{\boldsymbol{\theta}}^{\text{DY}}, \boldsymbol{\theta}_0 \rangle}{\| \boldsymbol{\theta}_0 \|_2^2} \, ,$$

for $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top$ and where $\hat{\boldsymbol{\theta}}^{\text{DY}}$ is the MDYPL estimator from $\mathcal{N}(\boldsymbol{0}_p, \boldsymbol{I}_p)$ covariates and signal $\boldsymbol{\theta}_0 = \boldsymbol{L}^\top \boldsymbol{\beta}_0$. Hence, immediately by Lemma E.3, $\tilde{\mu}_n \to \mu_*$ as $n \to \infty$ almost surely and the same applies to $\tilde{\sigma}_n$. Secondly, by (S78), $\frac{1}{\sqrt{p}} \sum_{j=1}^{p} |\tau_j \boldsymbol{\beta}_{0,j}| = o(1)$. Finally, recalling that $\tau_j \hat{\boldsymbol{\beta}}_j^{\text{DY}} = \hat{\boldsymbol{\theta}}_j^{\text{DY}}$, the Cauchy-Schwarz inequality yields

$$\frac{1}{\sqrt{p}} \sum_{j=1}^{p} \left| \tau_j \hat{\boldsymbol{\beta}}_j^{\text{DY}} \right| = \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \left| \hat{\boldsymbol{\theta}}_j^{\text{DY}} \right| \le \left\| \hat{\boldsymbol{\theta}}^{\text{DY}} \right\|_2 \, ,$$
(S80)

and by Lemma E.1 and Theorem 3.2, $\lim_{n \to \infty} \| \hat{\boldsymbol{\theta}}^{\text{DY}} \|_2 = \mathcal{O}(1)$ almost surely. Thus, (S79) is $o_p(1)$ as required. $\qquad\square$

Taken together, Lemma D.2 and (S74) imply that

$$\frac{1}{p} \sum_{j=1}^{p} \psi(\boldsymbol{T}_j) - \psi(\boldsymbol{Z}_j^{\text{scaled}}) = o_p(1) \, ,$$

so that it remains to characterise the distribution of $\boldsymbol{Z}^{\text{scaled}}$, which is fully characterised by the covariance structure $\boldsymbol{\Sigma}$. Zhao *et al.* (2022, Section B.2-B.4) show that if $\lambda_{\max}(\boldsymbol{\Sigma})/\lambda_{\min}(\boldsymbol{\Sigma}) = o\left(\sqrt{p}\right)$, then the empirical CDF of $\boldsymbol{Z}^{\text{scaled}}$ converges pointwise to that of a standard normal random variable in probability, which establishes the first part of the Theorem.

The proof of the second part of the Theorem follows a similar structure as the first part. In particular, for a bivariate pseudo-Lipschitz function ψ of order two, it is first shown that

$$\frac{1}{p} \sum_{j=1}^{p} \psi\left(\sigma_* \boldsymbol{T}_j, \tau_j \sqrt{n} \boldsymbol{\beta}_{0,j} \right) - \frac{1}{p} \sum_{j=1}^{p} \psi\left(\sigma_* \boldsymbol{T}_j^{\text{approx}}, \tau_j \sqrt{n} \boldsymbol{\beta}_{0,j} \right) = o_p(1) \, ,$$
(S81)

and that

$$\frac{1}{p} \sum_{j=1}^{p} \psi\left(\sigma_* \tilde{\boldsymbol{Z}}_j^{\text{scaled}}, \tau_j \sqrt{n} \boldsymbol{\beta}_{0,j} \right) - \frac{1}{p} \sum_{j=1}^{p} \psi\left(\sigma_* \boldsymbol{Z}_j^{\text{scaled}}, \tau_j \sqrt{n} \boldsymbol{\beta}_{0,j} \right) = o_p(1) \, ,$$
(S82)

which, in conjunction with (S74) yields

$$\frac{1}{p} \sum_{j=1}^{p} \psi\left(\sigma_* \boldsymbol{T}_j, \tau_j \sqrt{n} \boldsymbol{\beta}_{0,j} \right) - \frac{1}{p} \sum_{j=1}^{p} \psi\left(\sigma_* \boldsymbol{Z}_j^{\text{scaled}}, \tau_j \sqrt{n} \boldsymbol{\beta}_{0,j} \right) = o_p(1) \, ,$$

so that it then remains to analyse

$$\frac{1}{p} \sum_{j=1}^{p} \psi \left(\sigma_* \boldsymbol{Z}_j^{\mathrm{scaled}}, \tau_j \sqrt{n} \boldsymbol{\beta}_{0,j} \right) . \tag{S83}$$

Towards showing (S81), consider

$$\left| \frac{1}{p} \sum_{j=1}^{p} \psi \left(\sigma_* \boldsymbol{T}_j, \tau_j \sqrt{n} \boldsymbol{\beta}_{0,j} \right) - \psi \left(\sigma_* \boldsymbol{T}_j^{\mathrm{approx}}, \tau_j \sqrt{n} \boldsymbol{\beta}_{0,j} \right) \right|$$

$$\leq \frac{1}{p} \sum_{j=1}^{p} \left| \psi \left(\sigma_* \boldsymbol{T}_j, \tau_j \sqrt{n} \boldsymbol{\beta}_{0,j} \right) - \psi \left(\sigma_* \boldsymbol{T}_j^{\mathrm{approx}}, \tau_j \sqrt{n} \boldsymbol{\beta}_{0,j} \right) \right|$$

$$\leq \frac{L\sigma_*}{p} \sum_{j=1}^{p} \left(1 + \left\| [\sigma_* \boldsymbol{T}_j, \tau_j \sqrt{n} \boldsymbol{\beta}_{0,j}]^\top \right\|_2 + \left\| [\sigma_* \boldsymbol{T}_j^{\mathrm{approx}}, \tau_j \sqrt{n} \boldsymbol{\beta}_{0,j}]^\top \right\|_2 \right) \left| \boldsymbol{T}_j - \boldsymbol{T}_j^{\mathrm{approx}} \right|$$

$$\leq \frac{L\sigma_*}{p} \left\{ \sum_{j=1}^{p} \left(1 + \sigma_* |\boldsymbol{T}_j| + \sigma_* \left| \boldsymbol{T}_j^{\mathrm{approx}} \right| + 2\sqrt{n} \left| \tau_j \boldsymbol{\beta}_{0,j} \right| \right)^2 \right\}^{1/2} \left\{ \sum_{j=1}^{p} \left(\boldsymbol{T}_j - \boldsymbol{T}_j^{\mathrm{approx}} \right)^2 \right\}^{1/2} ,$$

where the last line follows by the Cauchy-Schwarz inequality and the monotonicity of ℓ_p norms. Now by construction of $\boldsymbol{T}_j, \boldsymbol{T}_j^{\mathrm{approx}}$ and since $(a+b)^2 \leq 2a^2 + 2b^2$,

$$\sum_{j=1}^{p} \left(\boldsymbol{T}_j - \boldsymbol{T}_j^{\mathrm{approx}} \right)^2 = \frac{n}{\sigma_*^2} \sum_{j=1}^{p} \left(\tau_j \hat{\boldsymbol{\beta}}_j^{\mathrm{DY}} \left(1 - \frac{\sqrt{\kappa}\sigma_*}{\tilde{\sigma}_n} \right) + \tau_j \boldsymbol{\beta}_{0,j} \left(\frac{\sqrt{\kappa}\sigma_*}{\tilde{\sigma}_n} \tilde{\mu}_n - \mu_* \right) \right)^2$$

$$\leq \frac{2n}{\sigma_*^2} \left\{ \left(1 - \frac{\sqrt{\kappa}\sigma_*}{\tilde{\sigma}_n} \right)^2 \sum_{j=1}^{p} \left(\tau_j \hat{\boldsymbol{\beta}}_j^{\mathrm{DY}} \right)^2 + \left(\frac{\sqrt{\kappa}\sigma_*}{\tilde{\sigma}_n} \tilde{\mu}_n - \mu_* \right)^2 \sum_{j=1}^{p} (\tau_j \boldsymbol{\beta}_{0,j})^2 \right\}$$

$$= \frac{2n}{\sigma_*^2} \left\{ \left(1 - \frac{\sqrt{\kappa}\sigma_*}{\tilde{\sigma}_n} \right)^2 \left\| \hat{\boldsymbol{\theta}}^{\mathrm{DY}} \right\|_2^2 + \left(\frac{\sqrt{\kappa}\sigma_*}{\tilde{\sigma}_n} \tilde{\mu}_n - \mu_* \right)^2 \left\| \boldsymbol{\theta}_0 \right\|_2^2 \right\}$$

$$= o(n) ,$$

by considerations similar to those in (S78)-(S80). Similarly, one can bound

$$\sum_{j=1}^{p} \left(1 + \sigma_* |\boldsymbol{T}_j| + \sigma_* \left| \boldsymbol{T}_j^{\mathrm{approx}} \right| + 2\sqrt{n} \left| \tau_j \boldsymbol{\beta}_{0,j} \right| \right)^2 \leq 4 \sum_{j=1}^{p} 1 + \sigma_*^2 \boldsymbol{T}_j^2 + \sigma_*^2 \boldsymbol{T}_j^{\mathrm{approx}\, 2} + 4n \tau_j^2 \boldsymbol{\beta}_{0,j}^2$$

$$= 4n + 4\sigma_*^2 \left\| \boldsymbol{T} \right\|_2^2 + 4\sigma_*^2 \left\| \boldsymbol{T}^{\mathrm{approx}} \right\|_2^2 + 16n \left\| \boldsymbol{\theta}_0 \right\|_2^2 .$$

Now $\boldsymbol{T}_j = \sqrt{n}(\hat{\boldsymbol{\theta}}_j^{\mathrm{DY}} - \mu_* \boldsymbol{\theta}_{0,j})/\sigma_*$ and by considerations analogous to Lemma E.3 and Theorem 3.1, choosing $\psi(t, u) = t^2$, yields that $\left\| \boldsymbol{T} \right\|_2^2 = \mathcal{O}_p(n)$. Similarly, by Lemma E.3 and the expansion of (S87), also $\left\| \boldsymbol{T}^{\mathrm{approx}} \right\|_2^2 = \mathcal{O}_p(n)$. Therefore (S81) is indeed $o_p(1)$.

Next, consider the difference (S82). Similarly to above,

$$\left| \frac{1}{p} \sum_{j=1}^{p} \psi \left(\sigma_* \tilde{\boldsymbol{Z}}_j^{\mathrm{scaled}}, \tau_j \sqrt{n} \boldsymbol{\beta}_{0,j} \right) - \psi \left(\sigma_* \boldsymbol{Z}_j^{\mathrm{scaled}}, \tau_j \sqrt{n} \boldsymbol{\beta}_{0,j} \right) \right|$$

$$\leq \frac{L\sigma_*}{p} \left\{ \sum_{j=1}^{p} \left(1 + \left| \boldsymbol{Z}_j^{\mathrm{scaled}} \right| + \left| \tilde{\boldsymbol{Z}}_j^{\mathrm{scaled}} \right| + 2\sqrt{n} \left| \tau_h \boldsymbol{\beta}_{0,j} \right| \right)^2 \right\}^{1/2} \left\{ \sum_{j=1}^{p} \left(\boldsymbol{Z}_j^{\mathrm{scaled}} - \tilde{\boldsymbol{Z}}_j^{\mathrm{scaled}} \right)^2 \right\}^{1/2} .$$

The term

$$\sum_{j=1}^{p} \left(1 + \left| \boldsymbol{Z}_j^{\mathrm{scaled}} \right| + \left| \tilde{\boldsymbol{Z}}_j^{\mathrm{scaled}} \right| + 2\sqrt{n} \left| \tau_j \boldsymbol{\beta}_{0,j} \right| \right)^2 = \mathcal{O}_p(n) ,$$

39

since $\boldsymbol{T}^{\mathrm{approx}} \overset{d}{=} \tilde{\boldsymbol{Z}}^{\mathrm{scaled}}$, which was shown to be bounded in the above paragraph and $\boldsymbol{Z}^{\mathrm{scaled}}$ was appropriately bounded in (S76)-(S77). On the other hand,

$$
\begin{aligned}
\sum_{j=1}^{p} \left(\boldsymbol{Z}_j^{\mathrm{scaled}} - \tilde{\boldsymbol{Z}}_j^{\mathrm{scaled}} \right)^2 &= \sum_{j=1}^{p} \left(\boldsymbol{Z}_j^{\mathrm{scaled}} \left(1 - \frac{\sqrt{p}}{\|\boldsymbol{\xi}\|_{\boldsymbol{\Sigma}}} \right) + \frac{\langle \boldsymbol{Z}, \boldsymbol{\beta}_0/\gamma_n \rangle_{\boldsymbol{\Sigma}}}{\gamma_n} \tau_j \boldsymbol{\beta}_{0,j} \right)^2 \\
&\leq 2 \left(1 - \frac{\sqrt{p}}{\|\boldsymbol{\xi}\|_{\boldsymbol{\Sigma}}} \right)^2 \|\boldsymbol{Z}^{\mathrm{scaled}}\|_2^2 + \left(\frac{\langle \boldsymbol{Z}, \boldsymbol{\beta}_0/\gamma_n \rangle_{\boldsymbol{\Sigma}}}{\gamma_n} \right)^2 \left(\frac{\sqrt{p}}{\|\boldsymbol{\xi}\|_{\boldsymbol{\Sigma}}} \right)^2 \|\boldsymbol{\theta}_0\|_2^2 \\
&= o_p(n),
\end{aligned}
$$

since arguments similar to Lemma G.10 show that $\lim_{n\to\infty} \sqrt{p}/\|\boldsymbol{\xi}\|_{\boldsymbol{\Sigma}} = 1$ almost surely, $\mathrm{E}[\boldsymbol{Z}_j^{\mathrm{scaled}\,2}] = 1$ and thus a Markov bound similar to (S76)-(S77) establishes that $\|\boldsymbol{Z}^{\mathrm{scaled}}\|_2^2 = \mathcal{O}_p(p)$. The second term is bounded since additionally $\lim_{n\to\infty} \gamma_n^2 = \lim_{n\to\infty} \|\boldsymbol{\theta}_0\|_2^2 = \gamma^2$, and $\langle \boldsymbol{Z}, \boldsymbol{\beta}_0/\gamma_n \rangle_{\boldsymbol{\Sigma}}$ is a normal random variable with $\mathcal{O}(1)$ variance.

Hence, it remains to analyse the behaviour of (S83) which is done in Zhao *et al.* (2022, Lemma C.1). As $\boldsymbol{Z}^{\mathrm{scaled}}$ solely depends on $\boldsymbol{\Sigma}$, this result directly applies to this setting. The conditions of said Lemma are satisfied in the current setting. Zhao *et al.* (2022, Lemma C.1) establishes that

$$
\frac{1}{p} \sum_{j=1}^{p} \psi \left(\sigma_* \boldsymbol{Z}_j^{\mathrm{scaled}}, \tau_j \sqrt{n} \boldsymbol{\beta}_{0,j} \right) \overset{p}{\longrightarrow} \mathrm{E} \left[\psi(\sigma_* G, \bar{\beta}) \right],
$$

for $G \sim \mathcal{N}(0,1) \perp \bar{\beta}$. This concludes the proof the Theorem.

E   Proof of Theorem 3.4

Theorem 3.4. *Assume that (α, κ, γ) are such that the conditions of Theorem 3.1 are met. Let $\hat{\boldsymbol{\beta}}^{DY}$ be the MDYPL estimator for a logistic regression model with i.i.d. $\mathcal{N}(\boldsymbol{0}_p, \boldsymbol{\Sigma})$ covariates and signal $\boldsymbol{\beta}_0$. Then for any regression coefficient such that $\sqrt{n}\tau_j \boldsymbol{\beta}_{0,j} = \mathcal{O}(1)$,*

$$
\sqrt{n}\tau_j \frac{\hat{\boldsymbol{\beta}}_j^{DY} - \mu_* \boldsymbol{\beta}_{0,j}}{\sigma_*} \overset{d}{\longrightarrow} \mathcal{N}(0,1),
$$

where $\tau_j^2 = var(\boldsymbol{x}_{ij}|\boldsymbol{x}_{i-j})$.

The proof of Theorem 3.4 is a replication of the arguments in the proof of Theorem 3.1 in Zhao *et al.* (2022) with the substitutions that are outlined in F and follows the general outline given in Section 3.2 of the main paper of this text. No claim of originality is made.

Lemma E.1. *For any matrix $\boldsymbol{L} \in \Re^{p\times p}$ and $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top$, let $\hat{\boldsymbol{\beta}}^{DY}$ be the MDYPL estimator from a logistic regression model with i.i.d. $\mathcal{N}(\boldsymbol{0}_p, \boldsymbol{\Sigma})$ covariates and signal $\boldsymbol{\beta}_0$. Then*

$$
\hat{\boldsymbol{\theta}}^{DY} = \boldsymbol{L}^\top \hat{\boldsymbol{\beta}}^{DY}
$$

is the MDYPL estimator from a logistic regression with $\mathcal{N}(\boldsymbol{0}_p, \boldsymbol{I}_p)$ covariates and signal $\boldsymbol{\theta}_0 = \boldsymbol{L}^\top \boldsymbol{\beta}_0$.

Proof. This result follows from the invariance of the MDYPL estimator under linear transformations of the covariates (see for example Zehna (1966)) and the rotational invariance of the normal distribution. Details can be found in Zhao *et al.* 2022, Proposition 2.1. $\square$

Lemma E.2. *Let $\hat{\boldsymbol{\theta}}^{DY}$ be the MDYPL estimator from a logistic regression model with i.i.d. $\mathcal{N}(\mathbf{0}_p, \boldsymbol{I}_p)$ covariates and signal $\boldsymbol{\theta}_0$. Define the random variables*

$$\mu_n = \frac{\langle \hat{\boldsymbol{\theta}}^{DY}, \boldsymbol{\theta}_0 \rangle}{\|\boldsymbol{\theta}_0\|_2^2}, \quad \sigma_n = \left\| \boldsymbol{P}_{\boldsymbol{\theta}_0^{\perp}} \hat{\boldsymbol{\theta}}^{DY} \right\|_2, \tag{S84}$$

where $\boldsymbol{P}_{\boldsymbol{\theta}_0^{\perp}}$ is the projection onto $\boldsymbol{\theta}_0^{\perp}$, the orthogonal complement of $\boldsymbol{\theta}_0$. Then

$$\frac{\hat{\boldsymbol{\theta}}^{DY} - \mu_n \boldsymbol{\theta}_0}{\sigma_n}$$

is uniformly distributed on the unit sphere lying in $\boldsymbol{\theta}_0^{\perp}$.

Proof. The proof is similar to Zhao *et al.* (2022, Lemma 2.1) and therefore omitted. □

Lemma E.3 (Zhao *et al.* (2022), Lemma 3.1). *Let $\hat{\boldsymbol{\beta}}^{DY}$ be the MDYPL estimator with i.i.d. covariates $\boldsymbol{x}_i \sim \mathcal{N}(\mathbf{0}_p, n^{-1}\boldsymbol{I}_p)$ and assume that the conditions of Theorem 3.1 are met. Then for μ_n, σ_n as in (S84),*

$$\mu_n \xrightarrow{a.s.} \mu_*, \quad \sigma_n \xrightarrow{a.s.} \sqrt{\kappa}\sigma_*,$$

where (μ_*, σ_*) are the solutions corresponding to the system of equations (S3) - (S5).

Proof. The proof is similar to the proof of Lemma 3.1 in Zhao *et al.* (2022) and relies on Lemma E.1 and Theorem 3.1.

Given $\boldsymbol{\theta}_0$ such that $\lim_{n \to \infty} \|\boldsymbol{\theta}_0\|_2 = \gamma$, let $\hat{\boldsymbol{\beta}}^{DY}$ be the MDYPL estimator from $\mathcal{N}(\mathbf{0}_p, n^{-1}\boldsymbol{I}_p)$ covariates and signal $\boldsymbol{\beta}_0 = \sqrt{n}\boldsymbol{U}\boldsymbol{\theta}_0$ where $\boldsymbol{U}$ is an orthogonal matrix such that $\boldsymbol{U}\boldsymbol{\theta}_0 = (\|\boldsymbol{\theta}_0\|_2, \ldots, \|\boldsymbol{\theta}_0\|_2)^{\top}/\sqrt{p}$, which can for example be constructed using a Householder transformation. Then by Lemma E.1, $\hat{\boldsymbol{\beta}}^{DY} = \sqrt{n}\boldsymbol{U}\hat{\boldsymbol{\theta}}^{DY}$ where $\hat{\boldsymbol{\theta}}^{DY}$ is the MDYPL estimator with $\mathcal{N}(\mathbf{0}_p, \boldsymbol{I}_p)$ covariates and signal $\boldsymbol{\theta}_0$. Note that by construction, the empirical distribution of $\boldsymbol{\beta}_0$ converges to a point mass at $\gamma/\sqrt{\kappa}$ and that $\lim_{n \to \infty} \frac{1}{p} \sum_{j=1}^n \beta_{0,j}^2 = \gamma^2/\kappa$ as required. Hence, Theorem 3.1 gives that for any pseudo-Lipschitz function ψ of order two,

$$\frac{1}{p} \sum_{j=1}^p \psi(\sqrt{n}([\boldsymbol{U}\hat{\boldsymbol{\theta}}^{DY}]_j - \mu_*[\boldsymbol{U}\boldsymbol{\theta}_0]_j), \sqrt{n}[\boldsymbol{U}\boldsymbol{\theta}_0]_j) \xrightarrow{a.s.} \mathrm{E}[\psi(\sigma_* G, \bar{\beta})],$$

where $G \sim \mathcal{N}(0,1)$, independent of $\bar{\beta} \sim \delta_{\gamma/\sqrt{\kappa}}$ and μ_*, σ_* solve (S3)-(S5). Then choosing $\psi(x, y) = xy$, gives that

$$\frac{1}{\kappa}\langle \boldsymbol{U}\hat{\boldsymbol{\theta}}^{DY} - \mu_*\boldsymbol{U}\boldsymbol{\theta}_0, \boldsymbol{U}\boldsymbol{\theta}_0 \rangle = \frac{1}{\kappa}\langle \hat{\boldsymbol{\theta}}^{DY} - \mu_*\boldsymbol{\theta}_0, \boldsymbol{\theta}_0 \rangle \xrightarrow{a.s.} 0,$$

from which it follows that

$$\mu_n = \frac{\langle \hat{\boldsymbol{\theta}}^{DY}, \boldsymbol{\theta}_0 \rangle}{\|\boldsymbol{\theta}_0\|_2} \xrightarrow{a.s.} \mu_*. \tag{S85}$$

Similarly, choosing $\psi(x, y) = x^2$, gives

$$\frac{1}{\kappa}\langle \boldsymbol{U}\hat{\boldsymbol{\theta}}^{DY} - \mu_*\boldsymbol{U}\boldsymbol{\theta}_0, \boldsymbol{U}\hat{\boldsymbol{\theta}}^{DY} - \mu_*\boldsymbol{U}\boldsymbol{\theta}_0 \rangle = \frac{1}{\kappa}\langle \hat{\boldsymbol{\theta}}^{DY} - \mu_*\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}^{DY} - \mu_*\boldsymbol{\theta}_0 \rangle = \left\| \hat{\boldsymbol{\theta}}^{DY} - \mu_*\boldsymbol{\theta}_0 \right\|_2^2 \xrightarrow{a.s.} \sigma_*^2,$$

which in conjunction with (S85) yields

$$\sigma_n^2 = \left\| \hat{\boldsymbol{\theta}}^{\mathrm{DY}} - \mu_n \boldsymbol{\theta}_0 \right\|_2^2 \xrightarrow{a.s.} \kappa \sigma_*^2$$

as required. □

Towards proving Theorem 3.4, note that Lemma E.1 implies that for any fixed coordinate j,

$$\sqrt{n} \tau_j \frac{\hat{\boldsymbol{\beta}}_j^{\mathrm{DY}} - \mu_* \boldsymbol{\beta}_{0,j}}{\sigma_*} = \sqrt{n} \frac{\hat{\boldsymbol{\theta}}_j^{\mathrm{DY}} - \mu_* \boldsymbol{\theta}_{0,j}}{\sigma_*} \,, \tag{S86}$$

where $\hat{\boldsymbol{\theta}}^{\mathrm{DY}} = \boldsymbol{L}^\top \hat{\boldsymbol{\beta}}^{\mathrm{DY}}$, is the MDYPL estimator from $\mathcal{N}(\boldsymbol{0}_p, \boldsymbol{I}_p)$ covariates with signal $\boldsymbol{\theta}_0 = \boldsymbol{L}^\top \boldsymbol{\beta}_0$ for $\boldsymbol{L}$ being the Cholesky factorization of $\boldsymbol{\Sigma} = \boldsymbol{L} \boldsymbol{L}^\top$ and where $\tau_j^2 = \mathrm{var}(\boldsymbol{x}_{ij} | \boldsymbol{x}_{i,-j})$ is the conditional variance of the jth covariate given all other covariates under $\boldsymbol{\Sigma}$ (see Zhao *et al.* (2022, Section 2.1)).

A simple expansion of the RHS of (S86) gives

$$\sqrt{n} \frac{\hat{\boldsymbol{\theta}}_j^{\mathrm{DY}} - \mu_* \boldsymbol{\theta}_{0,j}}{\sigma_*} = \sqrt{n} \frac{\hat{\boldsymbol{\theta}}_j^{\mathrm{DY}} - \mu_n \boldsymbol{\theta}_{0,j}}{\sigma_n} \frac{\sigma_n}{\sigma_*} + \sqrt{n} \frac{(\mu_n - \mu_*) \boldsymbol{\theta}_{0,j}}{\sigma_*} \,. \tag{S87}$$

Now by the assumption of the Theorem, $\sqrt{n} \boldsymbol{\theta}_{0,j} = \sqrt{n} \tau_j \boldsymbol{\beta}_{0,j} = \mathcal{O}(1)$. Hence, by Lemma E.3, the second term on the RHS of (S87) is $o(1)$ almost surely. Now by Lemma E.2, $(\hat{\boldsymbol{\theta}}_j^{\mathrm{DY}} - \mu_n \boldsymbol{\theta}_{0,j})/\sigma_n = \boldsymbol{P}_{\boldsymbol{\theta}_0^\perp} \hat{\boldsymbol{\theta}}^{\mathrm{DY}} / \| \boldsymbol{P}_{\boldsymbol{\theta}_0^\perp} \hat{\boldsymbol{\theta}}^{\mathrm{DY}} \|_2$ is distributed uniformly on the unit sphere lying in $\boldsymbol{\theta}_0^\perp$, where $\boldsymbol{P}_{\boldsymbol{\theta}_0^\perp}$ is the projection matrix into the orthogonal complement $\boldsymbol{\theta}_0^\perp$ of $\boldsymbol{\theta}_0$. It is a well known property of the Gaussian distribution that for $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}_p, \boldsymbol{I}_p)$, the projection onto the unit sphere is uniformly distributed (see for example Vershynin (2018), Section 3.3.3), and thus by Lemma E.2,

$$\frac{\boldsymbol{P}_{\boldsymbol{\theta}_0^\perp} \hat{\boldsymbol{\theta}}^{\mathrm{DY}}}{\left\| \boldsymbol{P}_{\boldsymbol{\theta}_0^\perp} \hat{\boldsymbol{\theta}}^{\mathrm{DY}} \right\|_2} \stackrel{d}{=} \frac{\boldsymbol{P}_{\boldsymbol{\theta}_0^\perp} \boldsymbol{Z}}{\left\| \boldsymbol{P}_{\boldsymbol{\theta}_0^\perp} \boldsymbol{Z} \right\|_2} \,,$$

for $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}_p, \boldsymbol{I}_p)$. Expanding $\boldsymbol{P}_{\boldsymbol{\theta}_0^\perp} \boldsymbol{Z}$ as

$$\boldsymbol{P}_{\boldsymbol{\theta}_0^\perp} \boldsymbol{Z} = \boldsymbol{Z} - \left\langle \boldsymbol{Z}, \frac{\boldsymbol{\theta}_0}{\| \boldsymbol{\theta}_0 \|_2} \right\rangle \frac{\boldsymbol{\theta}_0}{\| \boldsymbol{\theta}_0 \|_2} \,,$$

whose jth component is thus given by

$$Z_j \left(1 - \frac{\boldsymbol{\theta}_{0,j}^2}{\| \boldsymbol{\theta}_0 \|_2^2} \right) + \xi \,,$$

where Z_j is the jth component of $\boldsymbol{Z}$ and ξ is a normal random variable independent of Z_j with mean zero and variance

$$\mathrm{var}(\xi) = \frac{\| \boldsymbol{\theta}_{0,-j} \|_2^2}{\| \boldsymbol{\theta}_0 \|_2^2} \frac{\boldsymbol{\theta}_{0,j}^2}{\| \boldsymbol{\theta}_0 \|_2^2} = \mathcal{O}(n^{-1}) \,,$$

since $\sqrt{n} \boldsymbol{\theta}_j = \mathcal{O}(1)$. Consequently $\xi = \mathcal{O}_p(n^{-1/2})$. Further, by Lemma G.10, $\| \boldsymbol{P}_{\boldsymbol{\theta}_0^\perp} \boldsymbol{Z} \|_2 / \sqrt{p} \to 1$ almost

surely. Hence, taken together, one gets that

$$
\begin{aligned}
\sqrt{n}\frac{\hat{\boldsymbol{\theta}}_j^{\mathrm{DY}} - \mu_* \boldsymbol{\theta}_{0,j}}{\sigma_*} &= \sqrt{n}\frac{\hat{\boldsymbol{\theta}}_j^{\mathrm{DY}} - \mu_n \boldsymbol{\theta}_{0,j}}{\sigma_n}\frac{\sigma_n}{\sigma_*} + \sqrt{n}\frac{(\mu_n - \mu_*)\boldsymbol{\theta}_{0,j}}{\sigma_*} \\
&= \sqrt{n}\frac{\hat{\boldsymbol{\theta}}_j^{\mathrm{DY}} - \mu_n \boldsymbol{\theta}_{0,j}}{\sigma_n}\frac{\sigma_n}{\sigma_*} + o_p(1) \\
&= \sqrt{n}\frac{\hat{\boldsymbol{\theta}}_j^{\mathrm{DY}} - \mu_n \boldsymbol{\theta}_{0,j}}{\sigma_n} + o_p(1) \\
&\stackrel{d}{=} \sqrt{\frac{n}{p}}\frac{\left\|\boldsymbol{P}_{\boldsymbol{\theta}_0^\perp}\boldsymbol{Z}\right\|_2}{\sqrt{p}}\left\{Z_j\left(1 - \frac{\boldsymbol{\theta}_{0,j}^2}{\|\boldsymbol{\theta}_0\|_2^2}\right) + \xi\right\} + o_p(1) \\
&= \frac{1}{\sqrt{\kappa}}\left\{Z_j\left(1 - \frac{\boldsymbol{\theta}_{0,j}^2}{\|\boldsymbol{\theta}_0\|_2^2}\right) + \xi\right\} + o_p(1) \\
&= \frac{1}{\sqrt{\kappa}}Z_j\sigma_j + o_p(1) \\
&= \frac{1}{\sqrt{\kappa}}Z_j + o_p(1)
\end{aligned}
$$

for $\sigma_j = 1 - \boldsymbol{\theta}_{0,j}^2/\|\boldsymbol{\theta}_0\|_2^2$ which converges to one since $\boldsymbol{\theta}_{0,j} = \mathcal{O}(n^{-1/2})$. By (S86) and (S87),

$$
\sqrt{n}\frac{\hat{\boldsymbol{\theta}}_j^{\mathrm{DY}} - \mu_* \boldsymbol{\theta}_{0,j}}{\sigma_*} \stackrel{d}{=} \frac{1}{\sqrt{\kappa}}Z_j + o_p(1)\,,
\tag{S88}
$$

and the first part of the Theorem follows. The general result follows from (S88) and the rotational invariance of the Gaussian distribution in conjunction with Lemma E.1. The argument is completely similar to the one provided in the proof of Theorem 3.1 in Zhao *et al.* (2022) starting below (A.6) and is thus omitted.

F   Proof of Theorem 3.5

Theorem 3.5. *Let $\hat{\boldsymbol{\beta}}^{\mathrm{DY}}$ be the MDYPL estimator for a logistic regression model with i.i.d. $\mathcal{N}(\boldsymbol{0}_p, \boldsymbol{\Sigma})$ covariates and signal $\boldsymbol{\beta}_0$. Assume that (α, κ, γ) are such that $\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\|_2 = \mathcal{O}(n^{1/2})$ almost surely and that (S3)-(S5) admit a solution (μ_*, b_*, σ_*) such that $\boldsymbol{J}(\mu_*, b_*, \sigma_*)$ is nonsingular. Consider a fixed subset of indices $I = \{i_1, \ldots, i_k\}$, and the DY prior penalized likelihood ratio test statistic*

$$
\Lambda_I = \max_{\boldsymbol{\beta} \in \Re^p} \ell(\boldsymbol{\beta}; \boldsymbol{y}^*, \boldsymbol{X}) - \max_{\substack{\boldsymbol{\beta} \in \Re^p: \\ \boldsymbol{\beta}_j = 0,\, j \in I}} \ell(\boldsymbol{\beta}; \boldsymbol{y}^*, \boldsymbol{X})\,.
$$

Then, under the null that $\boldsymbol{\beta}_{0,i_1} = \ldots = \boldsymbol{\beta}_{0,i_k} = 0$, Λ_I is asymptotically distributed as

$$
2\Lambda_I \xrightarrow{d} \frac{\kappa\sigma_*^2}{b_*}\chi_k^2\,,
$$

where χ_k^2 is a Chi-squared random variable with k degrees of freedom.

This follows immediately from Theorem F.1 and Lemma E.1 as in the discussion of Zhao *et al.* (2022, Section 4).

Theorem F.1. *Let $\hat{\boldsymbol{\beta}}^{\mathrm{DY}}$ be the MDYPL estimator for a logistic regression model with i.i.d. $\mathcal{N}(\boldsymbol{0}_p, n^{-1}\boldsymbol{I}_p)$ covariates and signal $\boldsymbol{\beta}_0$. Assume that (α, κ, γ) are such that $\|\hat{\boldsymbol{\beta}}^{\mathrm{DY}}\|_2 = \mathcal{O}(n^{1/2})$ almost surely and that (S3)-(S5) admit a solution (μ_*, b_*, σ_*) such that $\boldsymbol{J}(\mu_*, b_*, \sigma_*)$ is nonsingular. Consider a fixed subset of*

indices $I = \{i_1, \ldots, i_k\}$, and the DY prior penalized likelihood ratio test statistic

$$\Lambda_I = \max_{\boldsymbol{\beta} \in \mathfrak{R}^p} \ell(\boldsymbol{\beta}; \boldsymbol{y}^*, \boldsymbol{X}) - \max_{\substack{\boldsymbol{\beta} \in \mathfrak{R}^p: \\ \boldsymbol{\beta}_j = 0, \, j \in I}} \ell(\boldsymbol{\beta}; \boldsymbol{y}^*, \boldsymbol{X}).$$

Then, under the null that $\boldsymbol{\beta}_{0,i_1} = \ldots = \boldsymbol{\beta}_{0,i_k} = 0$, Λ_I is asymptotically distributed as

$$2\Lambda_I \xrightarrow{d} \frac{\kappa \sigma_*^2}{b_*} \chi_k^2,$$

where χ_k^2 is a Chi-squared random variable with k degrees of freedom.

The proof of Theorem F.1 is analogous the proof of Theorem 3 in Sur and Candès (2019) where the MLE is replaced with the MDYPL estimator, the state evolutions in (5) are replaced by (S3)-(S5) and the standard logistic regression log-likelihood $\ell(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{X})$ is replaced by the log-likelihood with pseudo responses $\boldsymbol{y}^*$, i.e. $\ell(\boldsymbol{\beta}; \boldsymbol{y}^*, \boldsymbol{X})$ and their Theorem 2 is replaced by Theorem 3.1. A line by line analysis reveals that the arguments in the proof of Sur and Candès (2019) go through with minuscule adjustments. As the proof is lengthy and technical, and there are no noteworthy contributions in the adaptation to the MDYPL estimator, it is omitted.

G   Auxiliary Lemmas

Lemma G.1. *Given $\epsilon > 0$, $\alpha \in [0,1]$, consider the approximation $h_\epsilon(Z, \bar{\varepsilon}) = \frac{\alpha}{2} \left\{ 1 - \frac{\bar{\varepsilon} - \zeta'(Z)}{|\bar{\varepsilon} - \zeta'(Z)| + \epsilon} \right\} + \frac{1-\alpha}{2}$ to $h(Z, \bar{\varepsilon}) = \frac{\alpha}{2} \left\{ 1 - sign\left(\bar{\varepsilon} - \zeta'(Z) \right) \right\} + \frac{1-\alpha}{2}$. Then*

i) $\lim_{\epsilon \to 0} h_\epsilon(Z, \bar{\varepsilon}) = h(Z, \bar{\varepsilon})$ for all $Z, \bar{\varepsilon}$

ii) $h_\epsilon(Z, \bar{\varepsilon})$ is Lipschitz in $[Z, \bar{\varepsilon}]^\top$

iii) $\frac{\partial}{\partial Z} h_\epsilon(Z, \bar{\varepsilon})$ exists for all $Z, \bar{\varepsilon} \in \mathfrak{R}$, $\epsilon > 0$ and is Lipschitz in $[Z, \bar{\varepsilon}]^\top$.

Proof. i) If $\bar{\varepsilon} - \zeta'(Z) = 0$, then $h_\epsilon(Z, \bar{\varepsilon}) = 0$ for all $\epsilon > 0$. If $\bar{\varepsilon} - \zeta'(Z) \neq 0$, then $\lim_{\epsilon \to 0} h_\epsilon(Z, \bar{\varepsilon}) = (\bar{\varepsilon} - \zeta'(Z))/|\bar{\varepsilon} - \zeta'(Z)| = \mathrm{sign}(\bar{\varepsilon} - \zeta'(Z))$. The claim follows.

ii) First it is shown that $f(x) = x/(|x| + \epsilon)$ is Lipschitz for any $\epsilon > 0$ with Lipschitz constant ϵ^{-1}. To that end, note that $\frac{\partial}{\partial x} f(x) = \epsilon/(|x| + \epsilon)^2 \leq \epsilon^{-1}$ and by the mean value theorem (e.g. Rudin 1976, Theorem 5.10),

$$|f(x) - f(y)| \leq \sup_{z \in \mathfrak{R}} \left| \frac{\partial}{\partial z} f(z) \right| |x - y| \leq \frac{1}{\epsilon} |x - y|.$$

Further, note that $\zeta''(x)$ is positive and bounded from above by $1/4$. Hence,

$$
\begin{aligned}
|h_\epsilon(Z, \bar{\varepsilon}) - h_\epsilon(Z', \bar{\varepsilon}')| &= \frac{\alpha}{2} |f(\bar{\varepsilon} - \zeta'(Z)) - f(\bar{\varepsilon}' - \zeta'(Z'))| \\
&\leq \frac{\alpha}{2\epsilon} |\bar{\varepsilon} - \zeta'(Z) - (\bar{\varepsilon}' - \zeta'(Z'))| \\
&\leq \frac{\alpha}{2\epsilon} \{|\bar{\varepsilon} - \bar{\varepsilon}'| + |\zeta'(Z) - \zeta'(Z')|\} \\
&\leq \frac{\alpha}{2\epsilon} \left\{|\bar{\varepsilon} - \bar{\varepsilon}'| + \frac{1}{4} |Z - Z'|\right\} \\
&\leq \frac{\alpha}{8\epsilon} \{|\bar{\varepsilon} - \bar{\varepsilon}'| + |Z - Z'|\} \\
&\leq \frac{\alpha\sqrt{2}}{8\epsilon} \left\| \begin{bmatrix} Z \\ \bar{\varepsilon} \end{bmatrix} - \begin{bmatrix} Z' \\ \bar{\varepsilon}' \end{bmatrix} \right\|_2,
\end{aligned}
$$

where the first inequality follows by Lipschitzianity of $f(\cdot)$, the second by the triangle inequality, the third by the Mean Value Theorem and since $\sup_{x \in \Re} |\zeta''(x)| \leq 1/4$, and the last since $(|a| + |b|)^2 \leq 2(|a|^2 + |b|^2)$.

iii) Existence follows by differentiability of $f(x)$ and the chain rule,

$$
\frac{\partial}{\partial Z} h_\epsilon(Z, \bar{\varepsilon}) = \frac{\alpha}{2} \frac{\epsilon}{(|\bar{\varepsilon} - \zeta'(Z)| + \epsilon)^2} \zeta''(Z). \tag{S89}
$$

Next, it is shown that $f'(x) = \frac{\partial}{\partial x} x/(|x| + \epsilon) = \epsilon/(|x| + \epsilon)^2$ is Lipschitz in x.

Case 1: $x, y > 0$: For $x > 0$, the function simplifies to

$$
f'(x) = \frac{\epsilon}{(x + \epsilon)^2}.
$$

Differentiating with respect to x yields

$$
f''(x) = \frac{-2\epsilon}{(x + \epsilon)^3},
$$

which exists for all $x > 0$, and it is continuous and monotonically increasing in x. Furthermore, as x approaches infinity, $f''(x)$ approaches 0, and as x approaches 0 from the right, $f'(x)$ approaches $-2/\epsilon^2$. Thus, $|f''(x)|$ is bounded by $\frac{2}{\epsilon^2}$ for all $x > 0$.

By the Mean Value Theorem it follows that

$$
|f'(x) - f'(y)| \leq \frac{2}{\epsilon^2} |x - y|.
$$

Case 2: $x > 0, y < 0$: Note that

$$
|f'(x) - f'(y)| = \left| \frac{\epsilon}{(x + \epsilon)^2} - \frac{\epsilon}{(-y + \epsilon)^2} \right| = \left| \frac{\epsilon}{(x + \epsilon)^2} - \frac{\epsilon}{(z + \epsilon)^2} \right| = |f'(x) - f'(z)|,
$$

for $z = -y > 0$. Thus analogous to Case 1,

$$
|f'(x) - f'(z)| \leq \frac{2}{\epsilon^2} |x - z| < \frac{2}{\epsilon^2} |x + z| = \frac{2}{\epsilon^2} |x - y|,
$$

since $z = -y > 0, x > 0$.

45

Case 3: $x > 0, y = 0$

The difference $|f'(x) - \epsilon^{-1}|$ simplifies to

$$\frac{x^2 + 2\varepsilon x}{x^2 + 2\epsilon x + \epsilon^2}.$$

This expression is less than Cx if and only if

$$x^2 + 2\epsilon x \leq Cx^3 + 2C\epsilon x^2 + C\epsilon^2 x.$$

A sufficient condition for this inequality is $C > 2/\epsilon$.

Case 4: $x, y < 0$

By symmetry, the function behaves the same way for $x, y < 0$ as it does for $x, y > 0$. Thus, the Lipschitz constant derived for $x, y > 0$ also applies here.

This establishes that $f'(x)$ is indeed Lipschitz in x for all $x \in \Re$. Claim (iii) follows since $\zeta''(x)$ is Lipschitz in x and $\bar{\varepsilon} - \zeta'(Z)$ is Lipschitz in both $Z, \bar{\varepsilon}$ and by (S89). $\qquad\square$

Lemma G.2. *The following facts hold for the proximal operator* $prox_{b\zeta}(x)$ *and the function* $\Psi(y, s) = b(y - prox_{b\zeta}(by + s))$:

i) *The proximal operator* $prox_{b\zeta}(x)$ *is nonexpansive, i.e. for* $x, x' \in \Re$:

$$\left| prox_{b\zeta}(x) - prox_{b\zeta}(x') \right| \leq |x - x'|.$$

ii) $\Psi(h_\epsilon(Z, \bar{\varepsilon}), s) = b(h_\epsilon(Z, \bar{\varepsilon}) - \zeta'(prox_{b\zeta}(bh_\epsilon(Z, \bar{\varepsilon}) + s)))$ *is Lipschitz continuous in* $[Z, s, \bar{\varepsilon}]^\top$.

iii) $\Psi(y, s) = b(y - \zeta'(prox_{b\zeta}(by + s)))$ *is Lipschitz continuous in* s *with Lipschitz constant at most* 1.

iv) $\frac{\partial}{\partial Z}\Psi(h_\epsilon(Z, \bar{\varepsilon}), s)$ *exists for all* $Z, \bar{\varepsilon}, s \in \Re$, $\epsilon > 0$ *and is Lipschitz*

Proof. i) Note that $\zeta(x)$ is proper, closed and convex (see e.g. Beck 2017, Chapter 2, for the definition). Indeed, properness of $\zeta(x) : \Re \to [\infty, \infty]$ follows since it is monotonically increasing in x, $\lim_{x \to \infty} \zeta(x) = 0$ and $f(0) = 1$. Closedness follows since lower semicontinuity is equivalent to closedness (e.g. Beck 2017, Theorem 2.6) and $\zeta(x)$ is clearly continuous and therefore closed. Hence the claim follows by nonexpansivity of the proximal operator for proper, closed and convex functions (see for example Beck 2017, Theorem 6.4.2)b)), for $x, x' \in \Re$,

$$\left| prox_{b\zeta}(x) - prox_{b\zeta}(x') \right| \leq |x - x'|.$$

ii) It is shown in Lemma G.1 ii) that the approximation h_ϵ is Lipschitz.

By nonexpansivity of the proximal operator, which is shown in Lemma G.2 i),

$$\left| prox_{b\zeta}(x) - prox_{b\zeta}(x') \right| \leq |x - x'|.$$

Now letting $x = bh_\epsilon(Z, \bar{\varepsilon}) + s$, $x' = bh_\epsilon(Z', \bar{\varepsilon}') + s'$ then by Lipschtzianity of h_ϵ, and the triangle inequality and Jensen's inequality, again

$$\begin{aligned}
\left| prox_{b\zeta}(bh_\epsilon(Z, \bar{\varepsilon}) + s) - prox_{b\zeta}(bh_\epsilon(Z', \bar{\varepsilon}') + s') \right| &\leq |b(h_\epsilon(Z, \bar{\varepsilon}) - h_\epsilon(Z', \bar{\varepsilon}')) + (s - s')| \\
&\leq bL\left\{|\bar{\varepsilon} - \bar{\varepsilon}'| + |Z - Z'|\right\} + |s - s'| \\
&\leq \max\left\{bL, 1\right\}\left\{|\bar{\varepsilon} - \bar{\varepsilon}'| + |Z - Z'| + |s - s'|\right\} \\
&\leq \sqrt{3}\max\left\{bL, 1\right\}\left\|[Z, \bar{\varepsilon}, s]^\top - [Z', \bar{\varepsilon}', s']^\top\right\|_2,
\end{aligned}$$

46

for some constant L. Hence, again by boundedness of $\frac{\partial^2}{\partial x^2}\zeta(x)$, it immediately follows that for any $t \in [q]$, $\mathbf{x}, \mathbf{x'y}, \mathbf{y'} \in \Re^q$, and $b > 0$, note that

$$|\Psi(h_\epsilon(Z,\bar{\varepsilon}),s) - \Psi(h_\epsilon(Z',\bar{\varepsilon}'),s')|$$
$$\leq b\,|h_\epsilon(Z,\bar{\varepsilon}) - h_\epsilon(Z,\bar{\varepsilon})|$$
$$+ b\,\left|\zeta'\left(\text{prox}_{b\zeta}\left(bh_\epsilon(Z,\bar{\varepsilon}) + s\right)\right) - \zeta'\left(\text{prox}_{b\zeta}\left(bh_\epsilon(Z',\bar{\varepsilon}') + s'\right)\right)\right|$$
$$\leq bL\left\{|\bar{\varepsilon} - \bar{\varepsilon}'| + |Z - Z'|\right\} +$$
$$\frac{b}{4}\max\{bL, 1\}\left\{|\bar{\varepsilon} - \bar{\varepsilon}'| + |Z - Z'| + |s - s'|\right\}$$
$$= \left(bL + \frac{b}{4}\max\{bL, 1\}\right)\left\{|\bar{\varepsilon} - \bar{\varepsilon}'| + |Z - Z'|\right\} + \frac{b}{4}\max\{bL, 1\}\,|s - s'|$$
$$\leq \sqrt{3}\left(bL + \frac{b}{4}\max\{bL, 1\}\right)\left\|[Z, s, \bar{\varepsilon}]^\top - Z', s', \bar{\varepsilon}']^\top\right\|_2\,.$$

iii) From Donoho and Montanari (2016), Proposition 6.3,

$$\frac{\partial}{\partial s}\Psi(y, s) = -\frac{b\zeta''(\text{prox}_{b\zeta}\left(by + s\right))}{1 + b\zeta''(\text{prox}_{b\zeta}\left(by + s\right))}\,, \tag{S90}$$

which is less than one in absolute value and the claim follows by the mean value theorem.

iv) By Proposition 6.3 of Donoho and Montanari (2016), the proximal operator $\text{prox}_{b\zeta}(x)$ is differentiable for $b > 0$, $x \in \Re$ with partial derivative

$$\frac{\partial}{\partial x}\text{prox}_{b\zeta}(x) = \left.\frac{1}{1 + b\zeta''(z)}\right|_{z = \text{prox}_{b\zeta}(x)}.$$

Consequently,

$$\frac{\partial}{\partial x}\zeta'(\text{prox}_{b\zeta}(x)) = \left.\frac{b\zeta''(z)}{1 + b\zeta''(z)}\right|_{z = \text{prox}_{b\zeta}(x)}\,,$$

and thus using Lemma G.1 iii), and the chain rule,

$$\frac{\partial}{\partial Z}\Psi(h_\epsilon(Z,\bar{\varepsilon}),s) = \left.\frac{\partial}{\partial y}\Psi(y, s)\right|_{y = h_\epsilon(Z,\bar{\varepsilon})}\frac{\partial}{\partial Z}h_\epsilon(Z,\bar{\varepsilon})$$
$$= b\left(1 - \frac{b\zeta''(\text{prox}_{b\zeta}\left(bh_\epsilon(Z,\bar{\varepsilon}) + s\right))}{1 + b\zeta''(\text{prox}_{b\zeta}\left(bh_\epsilon(Z,\bar{\varepsilon}) + s\right))}\right)\frac{\partial}{\partial Z}h_\epsilon(Z,\bar{\varepsilon})$$
$$= \frac{b}{1 + b\zeta''(\text{prox}_{b\zeta}\left(bh_\epsilon(Z,\bar{\varepsilon}) + s\right))}\frac{\partial}{\partial Z}h_\epsilon(Z,\bar{\varepsilon})$$

Thus by simple expansion and the triangle inequality,

$$
\left| \frac{\partial}{\partial Z} \Psi(h_\epsilon(Z, \bar\varepsilon), s) - \frac{\partial}{\partial Z} \Psi(h_\epsilon(Z', \bar\varepsilon'), s') \right|
$$

$$
= b \left| \left\{ \frac{1}{1 + b\zeta''(\mathrm{prox}_{b\zeta}(bh_\epsilon(Z, \bar\varepsilon) + s))} - \frac{1}{1 + b\zeta''(\mathrm{prox}_{b\zeta}(bh_\epsilon(Z', \bar\varepsilon') + s'))} \right\} \frac{\partial}{\partial Z} h_\epsilon(Z, \bar\varepsilon) \right.
$$

$$
\left. + \frac{1}{1 + b\zeta''(\mathrm{prox}_{b\zeta}(bh_\epsilon(Z', \bar\varepsilon') + s'))} \left\{ \frac{\partial}{\partial Z} h_\epsilon(Z', \bar\varepsilon') - \frac{\partial}{\partial Z} h_\epsilon(Z, \bar\varepsilon) \right\} \right|
$$

$$
\leq b \left| \frac{b}{1 + b\zeta''(\mathrm{prox}_{b\zeta}(bh_\epsilon(Z, \bar\varepsilon) + s))} - \frac{1}{1 + b\zeta''(\mathrm{prox}_{b\zeta}(bh_\epsilon(Z', \bar\varepsilon') + s'))} \right| \left| \frac{\partial}{\partial Z} h_\epsilon(Z, \bar\varepsilon) \right|
$$

$$
+ b \left| \frac{1}{1 + b\zeta''(\mathrm{prox}_{b\zeta}(bh_\epsilon(Z', \bar\varepsilon') + s'))} \right| \left| \frac{\partial}{\partial Z} h_\epsilon(Z', \bar\varepsilon') - \frac{\partial}{\partial Z} h_\epsilon(Z, \bar\varepsilon) \right|. \quad \text{(S91)}
$$

Now it is easily seen that the function $f(x) = 1/(1+x)$ is Lipschitz over the nonnegative real line as it is differentiable and its derivative $\frac{\partial}{\partial x} f(x) = -1/(1+x)^2$ is bounded from above by 1 in absolute value for $x \geq 0$. Further, it is easily seen that $\frac{\partial}{\partial Z} h_\epsilon(Z, \bar\varepsilon)$ is bounded in absolute value. Hence, using boundedness of $\zeta'''(x)$, Lemma G.1 ii) and Lemma G.2 i), it follows immediately that

$$
b \left| \frac{1}{1 + b\zeta''(\mathrm{prox}_{b\zeta}(bh_\epsilon(Z, \bar\varepsilon) + s))} - \frac{1}{1 + b\zeta''(\mathrm{prox}_{b\zeta}(bh_\epsilon(Z', \bar\varepsilon') + s'))} \right| \left| \frac{\partial}{\partial Z} h_\epsilon(Z, \bar\varepsilon) \right|
$$

$$
\leq bC \left| \frac{1}{1 + b\zeta''(\mathrm{prox}_{b\zeta}(bh_\epsilon(Z, \bar\varepsilon) + s))} - \frac{1}{1 + b\zeta''(\mathrm{prox}_{b\zeta}(bh_\epsilon(Z', \bar\varepsilon') + s'))} \right|
$$

$$
\leq b^2 CC' \left| \zeta''(\mathrm{prox}_{b\zeta}(bh_\epsilon(Z, \bar\varepsilon) + s)) - \zeta''(\mathrm{prox}_{b\zeta}(bh_\epsilon(Z', \bar\varepsilon') + s')) \right|
$$

$$
\leq b^2 CC'C'' \left| \mathrm{prox}_{b\zeta}(bh_\epsilon(Z, \bar\varepsilon) + s) - \mathrm{prox}_{b\zeta}(bh_\epsilon(Z', \bar\varepsilon') + s') \right|
$$

$$
\leq b^2 CC'C''C''' \left\| [Z, s, \bar\varepsilon]^\top - [Z', s', \bar\varepsilon']^\top \right\|,
$$

for some positive constants C, C', C'', C'''. Similarly, by nonnegativity of ζ'' and b, as well as Lemma G.1 iii),

$$
b \left| \frac{1}{1 + b\zeta''(\mathrm{prox}_{b\zeta}(bh_\epsilon(Z', \bar\varepsilon') + s'))} \right| \left| \frac{\partial}{\partial Z} h_\epsilon(Z', \bar\varepsilon') - \frac{\partial}{\partial Z} h_\epsilon(Z, \bar\varepsilon) \right|
$$

$$
\leq bC \left| \frac{\partial}{\partial Z} h_\epsilon(Z', \bar\varepsilon') - \frac{\partial}{\partial Z} h_\epsilon(Z, \bar\varepsilon) \right|
$$

$$
\leq bCC' \left\| [Z, s, \bar\varepsilon]^\top - [Z', s', \bar\varepsilon']^\top \right\|,
$$

for some positive constants C, C'. Along with (S91) this concludes the proof. $\qquad \square$

Lemma G.3. *Let $\boldsymbol{X} \in \Re^{n \times p}$ have i.i.d. entries $x_{ij} \sim \mathcal{N}\left(0, n^{-1}\right)$ and let $\lim_{n \to \infty} p/n = \kappa \in (0, 1)$. Then almost surely, $\|\boldsymbol{X}\|_2 < \infty$.*

Proof. This follows directly from Rudelson and Vershynin 2010, Theorem 2.1. $\qquad \square$

Lemma G.4. *For $i = 1, \ldots, n$, let $\bar\varepsilon_i$ be i.i.d. $U(0, 1)$ random variables and $\boldsymbol{\eta}_{0,i}$ be i.i.d. $N(0, \gamma^2)$ such that $\bar\varepsilon_i \perp \boldsymbol{\eta}_{0,j}$ for $i, j = 1, \ldots, n$. Further, let*

48

$$y_i^* = \alpha \mathbb{1}\left\{\bar{\varepsilon}_i < \zeta'(\boldsymbol{\eta}_{0,i})\right\} + \frac{1-\alpha}{2}, \quad {}_\epsilon y_i^* = \alpha \left\{1 - \frac{\bar{\varepsilon}_i - \zeta'(\boldsymbol{\eta}_{0,i})}{|\bar{\varepsilon}_i - \zeta'(\boldsymbol{\eta}_{0,i})| + \epsilon}\right\} + \frac{1-\alpha}{2}.$$

Then for $\boldsymbol{y}^* = (y_1^*, \ldots, y_n^*)$, ${}_\epsilon\boldsymbol{y}^* = ({}_\epsilon y_1^*, \ldots, {}_\epsilon y_n^*)$,

i) $\lim\limits_{n\to\infty} \frac{1}{p}\|\boldsymbol{y}^* - {}_\epsilon\boldsymbol{y}^*\|_2^2 = E\left[(y_i^* - {}_\epsilon y_i^*)^2\right]$ *almost surely,*

ii) $\lim\limits_{\epsilon\to 0} E\left[(y_i^* - {}_\epsilon y_i^*)^2\right] = 0$

Proof. (i) By definition $\frac{1}{p}\|\boldsymbol{y}^* - {}_\epsilon\boldsymbol{y}^*\|_2^2 = \frac{1}{p}\sum_{i=1}^n (y_i^* - {}_\epsilon y_i^*)^2$. By boundedness of $y_i^*, {}_\epsilon y_i^*$, $\mathrm{E}\left[(y_i^* - {}_\epsilon y_i^*)^2\right]$ exists and thus, by the strong law of large numbers (e.g. Vershynin 2018, Theorem 1.3.1), claim (i) follows. (ii) Note that since $\bar{\varepsilon}_i, \boldsymbol{\eta}_{0,i}$ are continuous independent random variables. Hence the event $\{\exists i : \bar{\varepsilon}_i = \zeta'(\boldsymbol{\eta}_{0,i})\}$ has measure zero. Hence, $\mathrm{E}[(y^* - {}_\epsilon y^*)^2] = \mathrm{E}[(h(\bar{\varepsilon}_i, \boldsymbol{\eta}_{0,i}) - {}_\epsilon h(\bar{\varepsilon}_i, \boldsymbol{\eta}_{0,i}))^2]$. The claim now follows from boundedness of $\mathrm{E}\left[(y_i^* - {}_\epsilon y_i^*)^2\right]$, the dominated convergence theorem and Lemma G.1 i). $\square$

Lemma G.5. *The map $f(\boldsymbol{\beta}) = \omega\left(\|\boldsymbol{\beta}\|_2/\sqrt{n}\right)$ of Lemma 7 in Sur (2019) is such that if $\lim\limits_{\epsilon\to 0}\|\boldsymbol{\beta}(\epsilon)\|_2 = r$, for some sequence $\boldsymbol{\beta}(\epsilon) \in \Re^p$ indexed by ϵ, then $\lim\limits_{\epsilon\to 0} f(\boldsymbol{\beta}(\epsilon)) = \omega\left(r/\sqrt{n}\right)$.*

Proof. It is readily checked that the function $\omega\left(\|\boldsymbol{\beta}\|_2/\sqrt{n}\right)$ is of the form $c_1 \inf\limits_{z:|z|\leq c_2\frac{\|\boldsymbol{\beta}\|_2}{\sqrt{n}}} \zeta''(z)$ for some constants $c_1, c_2 > 0$ that do not depend on $\boldsymbol{\beta}$. Now $\zeta''(z)$ is symmetric and strictly decreasing in $|z|$ so that $\inf\limits_{z:|z|\leq c_2\frac{\|\boldsymbol{\beta}\|_2}{\sqrt{n}}} \zeta''(z) = \zeta''(c_2\|\boldsymbol{\beta}\|_2/\sqrt{n})$. The claim follows by taking limits and the continuity of the function ζ''. $\square$

Lemma G.6. *Consider the system of equations $F(\mu, b, \sigma, \epsilon) = 0$ given by*

$$E\left[\frac{1}{1 + b\zeta''\left(prox_{b\zeta}\left(b_\epsilon Y^* + Z_*\right)\right)}\right] - 1 + \kappa = 0 \tag{S92}$$

$$\frac{b}{\kappa\gamma^2} E\left[Z\left\{{}_\epsilon Y^* - \zeta'\left(prox_{b\zeta}\left(b_\epsilon Y^* + Z_*\right)\right)\right\}\right] = 0 \tag{S93}$$

$$\frac{b^2}{\kappa^2} E\left[\left\{{}_\epsilon Y^* - \zeta'\left(prox_{b\zeta}\left(b_\epsilon Y^* + Z_*\right)\right)\right\}^2\right] - \sigma^2 = 0, \tag{S94}$$

where $Z \sim N(0, \gamma^2)$, $G \sim N(0,1)$, and $\bar{\varepsilon} \sim U(0,1)$, all independent of each other, and ${}_\epsilon Y^* = h_\epsilon(Z, \bar{\varepsilon})$, $Z_* = \mu Z + \kappa^{1/2}\sigma G$.

Assume there exists a solution $F(\mu_*, b_*, \sigma_*, 0) = 0$ such that the Jacobian with respect to (μ_*, b_*, σ_*) of $F(\mu_*, b_*, \sigma_*, 0)$ is nonsingular.

Then, in a neighbourhood around (μ_*, b_*, σ_*), $\epsilon = 0$, there exist unique functions ${}_\epsilon\mu_*, {}_\epsilon b_*, {}_\epsilon\sigma_*$ that are continuous in ϵ such that $F({}_\epsilon\mu_*, {}_\epsilon b_*, {}_\epsilon\sigma_*, \epsilon) = 0$.

Proof. The claim follows by the Implicit Function Theorem (e.g. Rudin 1976, Theorem 9.28), for which it is required that

(i) There exists a solution $F(\mu_*, b_*, \sigma_*, 0)$,

(ii) The Jacobian with respect to (μ_*, b_*, σ_*) of $F(\mu_*, b_*, \sigma_*, 0)$ is nonsingular

(iii) $F(\mu, b, \sigma, \epsilon)$ is continuously differentiable in a neighbourhood around (μ_*, b_*, σ_*), $\epsilon = 0$

(i) and (ii) are assumed so that it remains to show (iii). Hence, it is sufficient to show that the partial derivatives of (S92)-(S94) exist and are continuous in a neighbourhood around (μ_*, b_*, σ_*), $\epsilon = 0$. Begin with (S92) and consider

$$\frac{\partial}{\partial b} \frac{1}{1 + b\zeta''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)} = -\frac{1}{\left[1 + b\zeta''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)\right]^2} \frac{\partial}{\partial b} b\zeta''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)$$

$$= -\frac{1}{\left[1 + b\zeta''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)\right]^2} \left\{ \zeta''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \right.$$

$$\left. + b\zeta'''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \left[\left. \frac{\partial}{\partial u}\text{prox}_{b\zeta}(u)\right|_{u=b_\epsilon Y^* + Z_*} \frac{\partial}{\partial b}u + \left. \frac{\partial}{\partial b}\text{prox}_{b\zeta}(u)\right|_{u=b_\epsilon Y^* + Z_*} \right] \right\}$$

$$= -\frac{1}{\left[1 + b\zeta''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)\right]^2} \left\{ \zeta''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \right.$$

$$\left. + b\zeta'''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \left[\frac{\epsilon Y^*}{1 + b\zeta''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)} - \frac{\zeta'\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)}{1 + b\zeta''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)} \right] \right\}$$

$$= -\frac{1}{\left[1 + b\zeta''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)\right]^2}$$

$$\cdot \left\{ \zeta''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) + \frac{b\zeta'''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)}{1 + b\zeta''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)} \left[\epsilon Y^* - \zeta'\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \right] \right\},$$
(S95)

where the first and second equality follow by the chain rule and the third uses Proposition A.1 of Donoho and Montanari (2016). Now note that $\zeta'(x) \in (0,1)$, $\zeta''(x) \in (0, \frac{1}{4}]$, $\zeta'''(x) \in \left[-\frac{1}{6\sqrt{3}}, \frac{1}{6\sqrt{3}}\right]$ and $_\epsilon Y^* \in [\frac{1-\alpha}{2}, \frac{1+\alpha}{2}]$. Hence (S95) is bounded. Thus, by the Dominated Convergence Theorem and the Mean Value Theorem,

$$\frac{\partial}{\partial b}\text{E}\left[\frac{1}{1 + b\zeta''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)} \right] = \text{E}\left[\frac{\partial}{\partial b} \frac{1}{1 + b\zeta''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)} \right]. \qquad (S96)$$

And again by dominated convergence, (S96) is continuous if (S95) is. By Lemma G.7, $\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)$ is continuous in $(b, _\epsilon Y^*, Z_*)$. Since $_\epsilon Y^* = \frac{\alpha}{2}\left\{ 1 - \frac{\bar{\varepsilon} - \zeta'(Z)}{|\bar{\varepsilon} - \zeta'(Z)| + \epsilon} \right\}$ and $Z_* = \mu Z + \kappa^{1/2}\sigma G$, $\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)$ is continuous in $(\mu, b\sigma, \epsilon)$. The claim follows by continuity of $\zeta'(\cdot), \zeta''(\cdot), \zeta'''(\cdot)$ and since (S95) is the composition of continuous functions. The treatment of the other partial derivatives is similar and thus details are omitted.

Next consider,

$$\frac{\partial}{\partial \mu} \frac{1}{1 + b\zeta''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)} = -\frac{b\zeta'''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) Z}{\left[1 + b\zeta''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)\right]^3}, \qquad (S97)$$

which is again dominated by an integrable function and thus $\frac{\partial}{\partial \mu}\text{E}\left[\frac{1}{1 + b\zeta''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)} \right]$ is continuous in $(\mu, b, \sigma, \epsilon)$.

Similarly,

$$\frac{\partial}{\partial \sigma} \frac{1}{1 + b\zeta''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)} = -\frac{b\zeta'''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \kappa^{1/2}G}{\left[1 + b\zeta''\left(\text{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)\right]^3},$$

and the same conclusion as for (S97) holds for $\frac{\partial}{\partial \sigma} \mathrm{E} \left[\frac{1}{1 + b\zeta''\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)} \right]$.

Finally,

$$\frac{\partial}{\partial \epsilon} \frac{1}{1 + b\zeta''\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)} = -\frac{b\zeta'''\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)}{\left[1 + b\zeta''\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)\right]^3} \frac{\alpha}{2} \frac{\bar{\varepsilon} - \zeta'(Z)}{(|\bar{\varepsilon} - \zeta'(Z)| + \epsilon)^2}, \quad \text{(S98)}$$

which is bounded and continuous in $(b, \mu, \sigma, \epsilon)$ for $\bar{\varepsilon} \neq \zeta'(Z)$. Since the event $\bar{\varepsilon} = \zeta'(Z)$, has measure zero, as $\bar{\varepsilon} \sim U(0, 1)$ and $Z \sim \mathcal{N}(0, 1)$ $\bar{\varepsilon} \perp Z$, dominated convergence still applies and the claim holds.

Next, consider (S93), and differentiate the integrand with respect to b,

$$\frac{\partial}{\partial b} \frac{b}{\kappa\gamma^2} Z \left\{ _\epsilon Y^* - \zeta'\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \right\} = \frac{1}{\kappa\gamma^2} Z \left\{ _\epsilon Y^* - \zeta'\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \right\}$$

$$- Z \frac{\zeta''\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)}{1 + b\zeta''\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)} \left\{ _\epsilon Y - \zeta'\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \right\},$$

of which both terms on the RHS are dominated by an integrable function and which are continuous in $(\mu, b, \sigma, \epsilon)$. Thus $\frac{\partial}{\partial b} \frac{b}{\kappa\gamma^2} \mathrm{E} \left[Z \left\{ _\epsilon Y^* - \zeta'\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \right\} \right]$ exists and is continuous in $(\mu, b, \sigma, \epsilon)$.

Next, differentiate with respect to μ to get

$$\frac{\partial}{\partial \mu} \frac{b}{\kappa\gamma^2} Z \left\{ _\epsilon Y^* - \zeta'\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \right\} = -\frac{b}{\kappa\gamma^2} \frac{Z^2 \zeta''\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)}{1 + b\zeta''\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)},$$

which is again dominated by an integrable function and continuous in $(\mu, b, \sigma, \epsilon)$, so that

$$\frac{\partial}{\partial \mu} \frac{b}{\kappa\gamma^2} \mathrm{E} \left[Z \left\{ _\epsilon Y^* - \zeta'\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \right\} \right]$$

exists and is continuous in $(\mu, b, \sigma, \epsilon)$.

Differentiating with respect to σ yields

$$\frac{\partial}{\partial \mu} \frac{b}{\kappa\gamma^2} Z \left\{ _\epsilon Y^* - \zeta'\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \right\} = -\frac{b}{\kappa\gamma^2} \frac{Z\kappa^{1/2} G \zeta''\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)}{1 + b\zeta''\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)},$$

and the same conclusion holds for $\frac{\partial}{\partial \sigma} \frac{b}{\kappa\gamma^2} \mathrm{E} \left[Z \left\{ _\epsilon Y^* - \zeta'\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \right\} \right]$.

Finally, differentiating with respect to ϵ,

$$\frac{\partial}{\partial \mu} \frac{b}{\kappa\gamma^2} Z \left\{ _\epsilon Y^* - \zeta'\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \right\} = -\frac{b}{\kappa\gamma^2} Z \left\{ 1 - \frac{b\zeta''\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)}{1 + b\zeta''\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)} \right\} \frac{\alpha}{2} \frac{\bar{\varepsilon} - \zeta'(Z)}{(|\bar{\varepsilon} - \zeta'(Z)| + \epsilon)^2},$$

and arguing analogous to (S98) yields that $\frac{\partial}{\partial \epsilon} \frac{b}{\kappa\gamma^2} \mathrm{E} \left[Z \left\{ _\epsilon Y^* - \zeta'\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \right\} \right]$ exists and is continuous in $(\mu, b, \sigma, \epsilon)$.

Lastly, the same is repeated for (S94).

$$\frac{\partial}{\partial b} \frac{b^2}{\kappa^2} \left\{ _\epsilon Y^* - \zeta'\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \right\}^2 - \sigma^2 = \frac{2b}{\kappa^2} \left\{ _\epsilon Y^* - \zeta'\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \right\}^2 - 1$$

$$- \frac{2b^2}{\kappa^2} \left\{ _\epsilon Y^* - \zeta'\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \right\} \frac{\zeta''\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)}{1 + b\zeta''\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)}$$

$$\cdot \left\{ _\epsilon Y^* - \zeta'\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \right\},$$

$$\frac{\partial}{\partial \mu} \frac{b^2}{\kappa^2} \left\{ _\epsilon Y^* - \zeta'\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \right\}^2 - \sigma^2 = -\frac{2b^2}{\kappa^2} \left\{ _\epsilon Y^* - \zeta'\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right) \right\} \frac{Z\zeta''\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)}{1 + b\zeta''\left(\mathrm{prox}_{b\zeta}(b_\epsilon Y^* + Z_*)\right)},$$

$$\frac{\partial}{\partial \sigma} \frac{b^2}{\kappa^2} \left\{ {}_\epsilon Y^* - \zeta' \left(\mathrm{prox}_{b\zeta} \left(b_\epsilon Y^* + Z_* \right) \right) \right\}^2 - \sigma^2 = -\frac{2b^2}{\kappa^2} \left\{ {}_\epsilon Y^* - \zeta' \left(\mathrm{prox}_{b\zeta} \left(b_\epsilon Y^* + Z_* \right) \right) \right\} \frac{\kappa^{1/2} G \zeta'' \left(\mathrm{prox}_{b\zeta} \left(b_\epsilon Y^* + Z_* \right) \right)}{1 + b\zeta'' \left(\mathrm{prox}_{b\zeta} \left(b_\epsilon Y^* + Z_* \right) \right)}$$
$$- 2\sigma \,,$$

$$\frac{\partial}{\partial \epsilon} \frac{b^2}{\kappa^2} \left\{ {}_\epsilon Y^* - \zeta' \left(\mathrm{prox}_{b\zeta} \left(b_\epsilon Y^* + Z_* \right) \right) \right\}^2 - \sigma^2 = -\frac{2b^2}{\kappa^2} \left\{ {}_\epsilon Y^* - \zeta' \left(\mathrm{prox}_{b\zeta} \left(b_\epsilon Y^* + Z_* \right) \right) \right\}$$
$$\cdot \frac{1}{1 + b\zeta'' \left(\mathrm{prox}_{b\zeta} \left(b_\epsilon Y^* + Z_* \right) \right)} \frac{\alpha}{2} \frac{\bar{\varepsilon} - \zeta'(Z)}{(|\bar{\varepsilon} - \zeta'(Z)| + \epsilon)^2} \,,$$

all of which are dominated by some integrable function and continuous almost everywhere. Thus, again by dominated convergence, all partial derivates of $\frac{b^2}{\kappa^2} \mathrm{E}\left[\left\{ {}_\epsilon Y^* - \zeta' \left(\mathrm{prox}_{b\zeta} \left(b_\epsilon Y^* + Z_* \right) \right) \right\}^2 \right] - \sigma^2$ exist and are continuous. Thus, $F(\mu, b, \sigma, \epsilon)$ is continuously differentiable in a neighbourhood around $(\mu_*, b_*, \sigma, 0)$ and the Implicit Function Theorem establishes the claim of the Lemma. $\qquad \square$

Lemma G.7. *The proximal operator $\mathrm{prox}_{b\zeta}(by + z)$ is continuous in (b, y, z) for $b > 0$, $x, y \in \Re$ where $\zeta(x) = \log(1 + e^x)$.*

Proof. Recall that by definition of the proximal operator,

$$\mathrm{prox}_{b\zeta}(by + z) = \arg\min_{u \in \Re} \left\{ b\zeta(u) + \frac{1}{2}(u - [by + z])^2 \right\} \,,$$

and since $\zeta(\cdot)$ is differentiable and strictly convex, it is uniquely determined as the solution to the first order condition

$$b\zeta'(u) + u - by - z = 0 \,.$$

Now note that $F(u, b, y, z) = b\zeta'(u) + u - by - z$ is continuously differentiable in (u, b, y, z) and that $\frac{\partial}{\partial u} F(u, b, y, z) = b\zeta''(u) + 1 > 0$ for all (u, b, y, z). Hence, by the Implicit Function Theorem (e.g. Rudin 1976, Theorem 9.28), there exists a unique function $u(b', y', z')$ in a neighbourhood around (b, y, z) that is continuous in its arguments such that $F(u(b', y', z'), b', y', z') = 0$. Hence, the proximal operator is continuous in a neighbourhood around (b, y, z). Since the choice of (b, y, z) was arbitrary, it is continuous over its entire domain. $\qquad \square$

Lemma G.8. *In the setting of Corollary B.2.1 and assuming that a stationary solution $({}_\epsilon\mu_*, {}_\epsilon b_*, {}_\epsilon\sigma_*)$ to the state evolution (S37)-(S38) exists, the AMP iterates of (S39) and (S40) satisfy the Cauchy property*

$$\lim_{t \to \infty} \lim_{N \to \infty} \frac{1}{p} \left\| {}_\epsilon \boldsymbol{\nu}^{t+1} - {}_\epsilon \boldsymbol{\nu}^t \right\|_2^2 = 0$$

$$\lim_{t \to \infty} \lim_{N \to \infty} \frac{1}{n} \left\| {}_\epsilon \boldsymbol{R}^{t+1} - {}_\epsilon \boldsymbol{R}^t \right\|_2^2 = 0$$

almost surely.

Proof. The proof is a direct translation of the Lemmas 6.8-6.9 of Donoho and Montanari (2016) along the lines suggested by Sur and Candès 2019 in the proof of Theorem 7 of the supplementary material. In particular, isolating iterates ${}_\epsilon\boldsymbol{\nu}^{t-1}, {}_\epsilon\boldsymbol{\nu}^t$ in Theorem B.2, similar arguments to those in

Corollary B.2.1 yield that, almost surely

$$\lim_{N \to \infty} \frac{1}{p} \sum_{j=1}^p \psi \left(\begin{bmatrix} {}_\epsilon \boldsymbol{\nu}_j^{t-1} \\ {}_\epsilon \boldsymbol{\nu}_j^t \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) = \mathrm{E} \left[\psi \left(\begin{bmatrix} \boldsymbol{Z}_t^{2t} \\ \boldsymbol{Z}_{t-1}^{2t} \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) \right], \quad \begin{bmatrix} \boldsymbol{Z}_{t-1}^{2t} \\ \boldsymbol{Z}_t^{2t} \end{bmatrix} \sim N \left(\boldsymbol{0}_2, \begin{bmatrix} {}_\epsilon \tau_{t-1}^2 & {}_\epsilon \rho_{t-1,t} \\ {}_\epsilon \rho_{t-1,t} & {}_\epsilon \tau_t^2 \end{bmatrix} \right)$$

where

$$_\epsilon \tau_t^2 = \kappa^{-2} \mathrm{E} \left[\{ \Psi^{t-1}(h_\epsilon(\boldsymbol{Z}_1^{2t-1}, \bar{\varepsilon}), \boldsymbol{Z}_2^{2t-1}) \}^2 \right]$$
$$_\epsilon \rho_{t-1,t} = \kappa^{-2} \mathrm{E} \left[\Psi^{t-1}(h_\epsilon(\boldsymbol{Z}_1^{2t-1}, \bar{\varepsilon}), \boldsymbol{Z}_{t+1}^{2t-1}) \Psi^{t-2}(h_\epsilon(\boldsymbol{Z}_1^{2t-1}, \bar{\varepsilon}), \boldsymbol{Z}_t^{2t-1}) \right],$$

which in turn have covariance structure according to the state evolution (S8)-(S9). Now recall that given the covariance structure derived in Corollary B.2.1, it was argued that given initial conditions $\hat{\boldsymbol{\beta}}^0$,

$$_\epsilon \mu_0 = \frac{1}{\gamma^2} \lim_{n \to \infty} \frac{\langle \hat{\boldsymbol{\beta}}^0, \boldsymbol{\beta}_0 \rangle}{n}$$
$$\tag{S99}$$
$$_\epsilon \sigma_0^2 = \lim_{n \to \infty} \frac{\left\| \hat{\boldsymbol{\beta}}^0 - \mu_0 \boldsymbol{\beta}_0 \right\|_2^2}{p},$$

for all t, $_\epsilon \tau_t = {}_\epsilon \sigma_t$. Now, given a stationary solutions $({}_\epsilon \mu_*, {}_\epsilon b_{*\epsilon}, \sigma_*)$ to the state evolutions (S36)-(S38), choose the oracle initialiser $\hat{\boldsymbol{\beta}}^0 = {}_\epsilon \mu_* \boldsymbol{\beta}_0 + {}_\epsilon \sigma_* \boldsymbol{\xi}$, where $\boldsymbol{\xi} \sim N(\boldsymbol{0}_p, \boldsymbol{I}_p)$ which is independent of $\boldsymbol{\beta}_0$. Then the initial conditions of (S99) are

$$_\epsilon \mu_0 = \frac{1}{\gamma^2} \lim_{n \to \infty} \frac{\langle {}_\epsilon \mu_* \boldsymbol{\beta}_0 + {}_\epsilon \sigma_* \boldsymbol{\xi}, \boldsymbol{\beta}_0 \rangle}{n} \overset{a.s}{=} {}_\epsilon \mu_*$$

$$_\epsilon \sigma_0^2 = \lim_{n \to \infty} \frac{\| \sigma_* \boldsymbol{\xi} \|_2^2}{p} \overset{a.s}{=} {}_\epsilon \sigma_*^2.$$

Henceforth condition on the event that $_\epsilon \mu_0 = {}_\epsilon \mu_*$, $_\epsilon \sigma_0^2 = {}_\epsilon \sigma_*^2$. Then $_\epsilon \tau_{t-1}^2 = {}_\epsilon \tau_t^2 = {}_\epsilon \sigma_*^2$ and letting $\psi(\boldsymbol{u}, \boldsymbol{v}) = (\boldsymbol{u}_1 - \boldsymbol{u}_2)^2$, for $\boldsymbol{u} = [\boldsymbol{u}_1, \boldsymbol{u}_2]^\top$, one gets that almost surely,

$$\lim_{N \to \infty} \frac{1}{p} \left\| {}_\epsilon \boldsymbol{\nu}^t - {}_\epsilon \boldsymbol{\nu}^{t-1} \right\|_2^2 = 2({}_\epsilon \sigma_*^2 - {}_\epsilon \rho_{t,t-1}).$$

Hence, the goal is to show that $\lim_{t \to \infty} {}_\epsilon \rho_{t,t-1} = {}_\epsilon \sigma_*^2$.

Towards that end, following the proof of Theorem 4.1 of Donoho and Montanari (2016), define the doubly infinite matrix $\boldsymbol{\Gamma} = \{ \Gamma_{s,t} \}_{s,t \in \mathbb{N}_0}$ with entries

$$\Gamma_{s+1,t+1} = \frac{1}{\kappa^2} \mathrm{E} \left[\Psi^* (h_\epsilon(Z, \bar{\varepsilon}), {}_\epsilon Z_s) \Psi^* (h_\epsilon(Z, \bar{\varepsilon}), {}_\epsilon Z_t) \right]$$

where $\Psi^*(y, s) = {}_\epsilon b_* \left(y - \zeta' \left(\mathrm{prox}_{{}_\epsilon b_* \zeta} ({}_\epsilon b_* y + s) \right) \right)$ and

$$\begin{matrix} {}_\epsilon Z_s = {}_\epsilon \mu_* Z + \kappa^{1/2} {}_\epsilon \tilde{G}_s \\ {}_\epsilon Z_t = {}_\epsilon \mu_* Z + \kappa^{1/2} {}_\epsilon \tilde{G}_t \end{matrix}, \quad \begin{bmatrix} {}_\epsilon \tilde{G}_s \\ {}_\epsilon \tilde{G}_t \end{bmatrix} \sim N \left(\boldsymbol{0}_2, \begin{bmatrix} {}_\epsilon \sigma_*^2 & \Gamma_{s,t} \\ \Gamma_{s,t} & {}_\epsilon \sigma_*^2 \end{bmatrix} \right),$$

where $Z \sim N(0, \gamma^2), \bar{\varepsilon}$, are all pairwise independent and independent of $\tilde{G}_s, \tilde{G}_t$. Further, take the boundary conditions $\Gamma_{0,0} = {}_\epsilon \sigma_*^2$, $\Gamma_{t,0} = \Gamma_{0,t} = 0$ so that in particular $\Gamma_{t,t} = {}_\epsilon \sigma_*^2$.

It is then readily verified that for all t, s, $\Gamma_{t,s} = {}_\epsilon \rho_{t,s}$. Next, let $q_t = \Gamma_{t+1,t}/{}_\epsilon \sigma_*^2$. Then q_t is governed by the recursion

$$q_{t+1} = H(q_t)$$
$$H(q) = \frac{1}{\kappa^2 {}_\epsilon \sigma_*^2} \mathrm{E} \left[\Psi^* (h_\epsilon(Z, \bar{\varepsilon}), {}_\epsilon Z_1) \Psi^* (h_\epsilon(Z, \bar{\varepsilon}), {}_\epsilon Z_2) \right],$$

53

with

$$\begin{aligned}
{}_\epsilon Z_1 &= {}_\epsilon\mu_* Z + \kappa^{1/2}{}_\epsilon\sigma_* G_1 \\
{}_\epsilon Z_2 &= {}_\epsilon\mu_* Z + \kappa^{1/2}{}_\epsilon\sigma_* G_2
\end{aligned}, \qquad
\begin{bmatrix} G_1 \\ G_2 \end{bmatrix} \sim N\left(\mathbf{0}_2, \begin{bmatrix} 1 & q \\ q & 1 \end{bmatrix} \right) \perp Z \sim N(0, \gamma^2).$$

Next, it is proven that:

i) $H(1) = 1$

ii) $H(q)$ is increasing for $q \in [0,1]$,

iii) $H(q)$ is strictly convex for $q \in [0,1]$

Towards i) note that

$$H(1) = \frac{1}{\kappa^2 {}_\epsilon\sigma_*^2} \mathrm{E}\left[\Psi^*\left(h_\epsilon(Z, \bar\varepsilon), {}_\epsilon Z_1 \right)^2 \right],$$

for ${}_\epsilon Z_1 = {}_\epsilon\mu_* Z + \kappa^{1/2}{}_\epsilon\sigma_{*\epsilon}^2 G$, $G \sim N(0,1) \perp Z$. Thus $H(1) = 1$ since (S38) holds for $({}_\epsilon\mu_*, {}_\epsilon b_*, {}_\epsilon\sigma_*)$.

Next, given $\bar\varepsilon, Z$, let $g_{Z,\bar\varepsilon}(G) = \Psi^*(h_\epsilon(Z, \bar\varepsilon), {}_\epsilon\mu_* Z + \kappa^{1/2}{}_\epsilon\sigma_* G)$ and let

$$\mathcal{H}(q) = \mathrm{E}\left[g_{Z,\bar\varepsilon}(G_1) g_{Z,\bar\varepsilon}(G_2) \Big| \bar\varepsilon, Z \right], \qquad
\begin{bmatrix} G_1 \\ G_2 \end{bmatrix} \sim N\left(\mathbf{0}_2, \begin{bmatrix} 1 & q \\ q & 1 \end{bmatrix} \right). \tag{S100}$$

Then, by linearity of expectations and the tower property, it is sufficient to show that $\mathcal{H}(q)$ satisfies ii) and iii).

Following the argument in Donoho and Montanari 2016, Lemma 6.9, let $\{X_t\}_{t \geq 0}$ be the stationary Ornstein-Uhlenbeck process with covariance $\mathrm{E}(X_0 X_t) = e^{-t}$. Then

$$\mathcal{H}(q) = \mathrm{E}\left[g_{Z,\bar\varepsilon}(G_1) g_{Z,\bar\varepsilon}(G_2) \right] \Big|_{t = \log(1/q)},$$

where the expectation is taken with respect to the process $\{X_t\}_{t \geq 0}$. Then by the spectral representation for $t = \log(1/q)$,

$$\mathcal{H}(q) = \sum_{l=1}^{\infty} c_l^2 e^{-lt} = \sum_{l=1}^{\infty} c_l^2 q_l,,$$

and the claim follows since $c_l \neq 0$ at least once for $g_{Z,\bar\varepsilon}$ nonlinear.

Hence, it follows that $\lim_{t \to \infty} q_t = 1$ if and only if $H'(1) \leq 1$.

Further, it holds that

$$H'(1) = \frac{1}{\kappa} \mathrm{E}\left[\left\{ \frac{\partial}{\partial s} \Psi^*(h_\epsilon(\bar\varepsilon, Z), {}_\epsilon Z_1) \right\}^2 \right].$$

Indeed, again condition on $\bar\varepsilon, Z$ and consider $\frac{\partial}{\partial q}\mathcal{H}(q)$ as defined in (S100). For notational convenience, the conditioning in the expectation of $\mathcal{H}$ is suppressed below and $g_{Z,\bar\varepsilon}(G)$ is simply denoted by $g(G)$.

Then, since G_2 can be written as $qG_1 + \sqrt{1-q^2}\tilde{G}$ for $\tilde{G} \sim N(0,1)$ is independent of all other

random variables, it holds that

$$
\begin{aligned}
\frac{\partial}{\partial q}\mathcal{H}(q) &= \frac{\partial}{\partial q}\mathrm{E}\left[g(G_1)g(G_2)\right] \\
&= \mathrm{E}\left[\frac{\partial}{\partial q}g(G_1)g(G_2)\right] \\
&= \mathrm{E}\left[\frac{\partial}{\partial q}g(G_1)g(qG_1 + \sqrt{1-q^2}\tilde{G})\right] \\
&= \mathrm{E}\left[g(G_1)g'(qG_1 + \sqrt{1-q^2}\tilde{G})\kappa^{1/2}_{\epsilon}\sigma_*\left\{G_1 - \frac{q}{\sqrt{1-q^2}}\tilde{G}\right\}\right] \\
&= \mathrm{E}\left[g(G_1)g'(qG_1 + \sqrt{1-q^2}\tilde{G})\kappa^{1/2}_{\epsilon}\sigma_*G_1\right] - \frac{q}{\sqrt{1-q^2}}\mathrm{E}\left[g(G_1)g'(qG_1 + \sqrt{1-q^2}\tilde{G})\kappa^{1/2}_{\epsilon}\sigma_*\tilde{G}\right],
\end{aligned}
$$

(S101)

where the second inequality follows by dominated convergence. Now let $F(G_1, G_2) = g(G_1)g'(G_2)$. Then note that $\frac{\partial}{\partial G_1}F(G_1, G_2) = \kappa^{1/2}_{\epsilon}\sigma_*g'(G_1)g'(G_2)$ and $\frac{\partial}{\partial G_2}F(G_1, G_2) = \kappa^{1/2}_{\epsilon}\sigma_*g(G_1)g''(G_2)$. Hence, using Gaussian integration by parts, which can be derived as a consequence of Stein's Lemma (Stein, 1981, Lemma 1),

$$
\begin{aligned}
\mathrm{E}\left[g(G_1)g'(qG_1 + \sqrt{1-q^2}\tilde{G})\kappa^{1/2}_{\epsilon}\sigma_*G_1\right] &= \mathrm{E}\left[g(G_1)g'(G_2)\kappa^{1/2}_{\epsilon}\sigma_*G_1\right] \\
&= \kappa^{1/2}_{\epsilon}\sigma_*\mathrm{E}\left[G_1F(G_1, G_2)\right] \\
&= \kappa^{1/2}_{\epsilon}\sigma_*\left\{\mathrm{cov}(G_1, G_1)\mathrm{E}\left[\frac{\partial}{\partial G_1}F(G_1, G_2)\right]\right. \\
&\quad \left. + \mathrm{cov}(G_1, G_2)\mathrm{E}\left[\frac{\partial}{\partial G_2}F(G_1, G_2)\right]\right\} \\
&= \kappa_{\epsilon}\sigma^2_*\left\{\mathrm{E}\left[\frac{\partial}{\partial G_1}F(G_1, G_2)\right] + q\mathrm{E}\left[\frac{\partial}{\partial G_2}F(G_1, G_2)\right]\right\} \\
&= \kappa_{\epsilon}\sigma^2_*\left\{\mathrm{E}\left[g'(G_1)g'(G_2)\right] + q\mathrm{E}\left[\frac{\partial}{\partial G_2}g(G_1)g''(G_2)\right]\right\}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathrm{E}\left[g(G_1)g'(qG_1 + \sqrt{1-q^2}\tilde{G})\kappa^{1/2}_{\epsilon}\sigma_*\tilde{G}\right] &= \kappa^{1/2}_{\epsilon}\sigma_*\mathrm{E}\left[F(G_1, qG_1 + \sqrt{1-q^2}\tilde{G})\tilde{G}\right] \\
&= \kappa^{1/2}_{\epsilon}\sigma_*\left\{\mathrm{cov}(G_1, \tilde{G})\mathrm{E}\left[\frac{\partial}{\partial G_1}F(G_1, qG_1 + \sqrt{1-q^2}\tilde{G})\right]\right. \\
&\quad \left. + \mathrm{cov}(\tilde{G}, \tilde{G})\mathrm{E}\left[\frac{\partial}{\partial \tilde{G}}F(G_1, qG_1 + \sqrt{1-q^2}\tilde{G})\right]\right\} \\
&= \sqrt{1-q^2}\kappa_{\epsilon}\sigma^2_*\mathrm{E}\left[g(G_1)g''(qG_1 + \sqrt{1-q^2}\tilde{G})\right] \\
&= \sqrt{1-q^2}\kappa_{\epsilon}\sigma^2_*\mathrm{E}\left[g(G_1)g''(G_2)\right],
\end{aligned}
$$

(S102)

where it was used that $\mathrm{cov}(G_1, \tilde{G}) = 0$. Thus combining (S101)-(S102), it follows that

$$
\frac{\partial}{\partial q}\mathcal{H}(q) = \kappa_{\epsilon}\sigma^2_*\mathrm{E}\left[g'(G_1)g'(G_2)\right].
$$

Marginalizing over $Z, \bar{\varepsilon}$ then gives

$$
\begin{aligned}
\frac{\partial}{\partial q} H(q) &= \frac{1}{\kappa^2 {}_\epsilon \sigma_*^2} \frac{\partial}{\partial q} \mathrm{E}_{Z,\bar{\varepsilon}} \left[\mathrm{E} \left[g(G_1) g(G_2) \Big| Z, \bar{\varepsilon} \right] \right] \\
&= \frac{1}{\kappa^2 {}_\epsilon \sigma_*^2} \mathrm{E}_{Z,\bar{\varepsilon}} \left[\mathrm{E} \left[\frac{\partial}{\partial q} g(G_1) g(G_2) \Big| Z, \bar{\varepsilon} \right] \right] \\
&= \frac{1}{\kappa} \mathrm{E}_{Z,\bar{\varepsilon}} \left[\mathrm{E} \left[g'(G_1) g'(G_2) \Big| Z, \bar{\varepsilon} \right] \right] \\
&= \frac{1}{\kappa} \mathrm{E} \left[g'(G_1) g'(G_2) \right] \\
&= \frac{1}{\kappa} \mathrm{E} \left[g'(G_1) g'(q G_1 + \sqrt{1 - q^2} \tilde{G}) \right] ,
\end{aligned}
$$

and it thus follows that $H'(1) = \frac{1}{\kappa} \mathrm{E} \left[\left\{ \frac{\partial}{\partial s} \Psi^*(h_\epsilon(\bar{\varepsilon}, Z), {}_\epsilon Z_1) \right\}^2 \right]$ as required.

Now it is easily checked that $\frac{\partial}{\partial s} \Psi^* \in (0, 1)$ and thus

$$
H'(1) = \frac{1}{\kappa} \mathrm{E} \left[\left\{ \frac{\partial}{\partial s} \Psi^*(h_\epsilon(\bar{\varepsilon}, Z), {}_\epsilon Z_1) \right\}^2 \right] \leq \frac{1}{\kappa} \mathrm{E} \left[\frac{\partial}{\partial s} \Psi^*(h_\epsilon(\bar{\varepsilon}, Z), {}_\epsilon Z_1) \right] = 1
$$

since (S36) holds. Similar arguments show the Cauchy-property for $\frac{1}{n} \| {}_\epsilon \boldsymbol{R}^t - {}_\epsilon \boldsymbol{R}^{t-1} \|_2^2$. $\qquad\square$

Lemma G.9. *The solution to the system of equations* (S3)-(S5) *is equivalent to the stationary solution of* (S16)-(S18)

Proof. Begin by imposing stationarity for equations (S16)-(S18), so that they become

$$
\begin{aligned}
1 - \kappa &= \mathrm{E} \left[\frac{1}{1 + b_* \zeta'' \left(\mathrm{prox}_{b_* \zeta} (b_* Y^* + Z_*) \right)} \right] \\
0 &= \frac{b_*}{\kappa \gamma^2} \mathrm{E} \left[Z \left\{ Y^* - \zeta' \left(\mathrm{prox}_{b_* \zeta} (b_* Y^* + Z_*) \right) \right\} \right] \\
\sigma_*^2 &= \frac{b_*^2}{\kappa^2} \mathrm{E} \left[\left\{ Y^* - \zeta' \left(\mathrm{prox}_{b_* \zeta} (b_* Y^* + Z_*) \right) \right\}^2 \right] ,
\end{aligned}
$$

for independent $Z \sim N(0, \gamma^2)$, $G \sim N(0, 1)$, and $\bar{\varepsilon} \sim U(0, 1)$, and $Y^* = \alpha \mathbb{1} \left\{ \bar{\varepsilon} \leq \zeta'(Z) \right\} + \frac{1-\alpha}{2}$, $Z_* = \mu_* Z + \kappa^{1/2} \sigma_* G$. Hence

$$
\begin{aligned}
\mathrm{E} \left[\frac{1}{1 + b_* \zeta'' \left(\mathrm{prox}_{b_* \zeta} (b_* Y^* + Z_*) \right)} \right] &= \mathrm{E} \left[\frac{\zeta'(Z)}{1 + b_* \zeta'' \left(\mathrm{prox}_{b_* \zeta} \left(b_* \frac{1+\alpha}{2} + Z_* \right) \right)} \right] + \mathrm{E} \left[\frac{\zeta'(-Z)}{1 + b_* \zeta'' \left(\mathrm{prox}_{b_* \zeta} \left(b_* \frac{1-\alpha}{2} + Z_* \right) \right)} \right] \\
&= \mathrm{E} \left[\frac{\zeta'(Z)}{1 + b_* \zeta'' \left(\mathrm{prox}_{b_* \zeta} \left(b_* \frac{1+\alpha}{2} + Z_* \right) \right)} \right] + \mathrm{E} \left[\frac{\zeta'(-Z)}{1 + b_* \zeta'' \left(-\mathrm{prox}_{b_* \zeta} \left(b_* \frac{1-\alpha}{2} + Z_* \right) \right)} \right] \\
&= \mathrm{E} \left[\frac{\zeta'(Z)}{1 + b_* \zeta'' \left(\mathrm{prox}_{b_* \zeta} \left(b_* \frac{1+\alpha}{2} + Z_* \right) \right)} \right] + \mathrm{E} \left[\frac{\zeta'(-Z)}{1 + b_* \zeta'' \left(\mathrm{prox}_{b_* \zeta} \left(b_* \frac{1+\alpha}{2} - Z_* \right) \right)} \right] \\
&= 2 \mathrm{E} \left[\frac{\zeta'(Z)}{1 + b_* \zeta'' \left(\mathrm{prox}_{b_* \zeta} \left(b_* \frac{1+\alpha}{2} + Z_* \right) \right)} \right] ,
\end{aligned}
$$

where the first equality follows by conditioning on Y^*, the second follows since $\zeta''(x) = \zeta''(-x)$,

56

the fourth since $-\text{prox}_{b\zeta}(-u) = \text{prox}_{b\zeta}(u+b)$ (cf. p.18 in the supplementary material of Sur and Candès (2019)), which holds since the proximal operator $x^* = \text{prox}_{b\zeta}(-u)$ is the unique solution to $b\zeta'(x^*) + x^* + u = 0$ and similarly $y^* = \text{prox}_{b\zeta}(u+b)$ is the unique solution to $b\zeta'(y^*) + y^* - u - b = 0$. Hence, by substitution and using that $\zeta'(-x) = 1 - \zeta'(x)$, y^* is the unique solution to $b(\zeta'(y) - \zeta'(-x^*)) + y^* - x = 0$ so that $y^* = -x^*$. Therefore, letting $-u = b_*(1-\alpha)/2 + Z_*$, $u + b_* = b_*(1+\alpha)/2 - Z_*$ as required. The last equality follows by a simple change of variable and the symmetry of the normal distribution.

Similarly,

$$
\begin{aligned}
\mathrm{E}\left[Z\left\{Y^* - \zeta'\left(\text{prox}_{b_*\zeta}(b_*Y^* + Z_*)\right)\right\}\right] &= \mathrm{E}\left[\zeta'(Z)Z\left\{\frac{1+\alpha}{2} - \zeta'\left(\text{prox}_{b_*\zeta}\left(b_*\frac{1+\alpha}{2} + Z_*\right)\right)\right\}\right] \\
&+ \mathrm{E}\left[\zeta'(-Z)Z\left\{\frac{1-\alpha}{2} - \zeta'\left(\text{prox}_{b_*\zeta}\left(b_*\frac{1-\alpha}{2} + Z_*\right)\right)\right\}\right] \\
&= \mathrm{E}\left[\zeta'(Z)Z\left\{\frac{1+\alpha}{2} - \zeta'\left(\text{prox}_{b_*\zeta}\left(b_*\frac{1+\alpha}{2} + Z_*\right)\right)\right\}\right] \\
&+ \mathrm{E}\left[\zeta'(-Z)Z\left\{\frac{1-\alpha}{2} - \left(1 - \zeta'\left(-\text{prox}_{b_*\zeta}\left(b_*\frac{1-\alpha}{2} + Z_*\right)\right)\right)\right\}\right] \\
&= \mathrm{E}\left[\zeta'(Z)Z\left\{\frac{1+\alpha}{2} - \zeta'\left(\text{prox}_{b_*\zeta}\left(b_*\frac{1+\alpha}{2} + Z_*\right)\right)\right\}\right] \\
&- \mathrm{E}\left[\zeta'(-Z)Z\left\{\frac{1+\alpha}{2} - \zeta'\left(\text{prox}_{b_*\zeta}\left(b_*\frac{1+\alpha}{2} - Z_*\right)\right)\right\}\right] \\
&= 2\mathrm{E}\left[\zeta'(Z)Z\left\{\frac{1+\alpha}{2} - \zeta'\left(\text{prox}_{b_*\zeta}\left(b_*\frac{1+\alpha}{2} + Z_*\right)\right)\right\}\right],
\end{aligned}
$$

where again, the first equality comes from conditioning on Y^*, the second, from $\zeta'(-x) = 1 - \zeta'(x)$, the third by the symmetry of the proximal operator around b, i.e. $-\text{prox}_{b\zeta}(-u) = \text{prox}_{b\zeta}(u+b)$, and the last by change of variable and symmetry of the normal distribution.

And finally,

$$
\begin{aligned}
\mathrm{E}\left[\left\{Y^* - \zeta'\left(\text{prox}_{b_*\zeta}(b_*Y^* + Z_*)\right)\right\}^2\right] &= \mathrm{E}\left[\zeta'(Z)\left\{\frac{1+\alpha}{2} - \zeta'\left(\text{prox}_{b_*\zeta}\left(b_*\frac{1+\alpha}{2} + Z_*\right)\right)\right\}^2\right] \\
&+ \mathrm{E}\left[\zeta'(-Z)\left\{\frac{1-\alpha}{2} - \zeta'\left(\text{prox}_{b_*\zeta}\left(b_*\frac{1-\alpha}{2} + Z_*\right)\right)\right\}^2\right] \\
&= \mathrm{E}\left[\zeta'(Z)\left\{\frac{1+\alpha}{2} - \zeta'\left(\text{prox}_{b_*\zeta}\left(b_*\frac{1+\alpha}{2} + Z_*\right)\right)\right\}^2\right] \\
&+ \mathrm{E}\left[\zeta'(-Z)\left\{\frac{1+\alpha}{2} - \zeta'\left(-\text{prox}_{b_*\zeta}\left(b_*\frac{1-\alpha}{2} + Z_*\right)\right)\right\}^2\right] \\
&= \mathrm{E}\left[\zeta'(Z)\left\{\frac{1+\alpha}{2} - \zeta'\left(\text{prox}_{b_*\zeta}\left(b_*\frac{1+\alpha}{2} + Z_*\right)\right)\right\}^2\right] \\
&+ \mathrm{E}\left[\zeta'(-Z)\left\{\frac{1+\alpha}{2} - \zeta'\left(\text{prox}_{b_*\zeta}\left(b_*\frac{1+\alpha}{2} - Z_*\right)\right)\right\}^2\right] \\
&= 2\mathrm{E}\left[\zeta'(Z)\left\{\frac{1+\alpha}{2} - \zeta'\left(\text{prox}_{b_*\zeta}\left(b_*\frac{1+\alpha}{2} + Z_*\right)\right)\right\}^2\right].
\end{aligned}
$$

$\square$

Lemma G.10. *Let $\boldsymbol{u} \in \Re^p$ be a unit vector and $\boldsymbol{P}_{\boldsymbol{u}^\perp} = \boldsymbol{I}_p - \boldsymbol{u}\boldsymbol{u}^\top$ be the projection into the subspace $\boldsymbol{u}^\perp$ that is orthogonal to $\boldsymbol{u}$. Then for $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}_p, \boldsymbol{I}_p)$, almost surely,*

$$\lim_{p \to \infty} \frac{\|\boldsymbol{P}_{\boldsymbol{u}^\perp}\boldsymbol{Z}\|_2}{\sqrt{p}} = 1\,,$$

as $p \to \infty$.

Proof. Consider $\|\boldsymbol{P}_{\boldsymbol{u}^\perp}\boldsymbol{Z}\|_2^2 = \boldsymbol{Z}^\top \boldsymbol{P}_{\boldsymbol{u}^\perp}\boldsymbol{Z}$ since $\boldsymbol{P}_{\boldsymbol{u}^\perp}$ is idempotent and note that $\mathrm{tr}\left(\boldsymbol{P}_{\boldsymbol{u}^\perp}\right) = p - 1$. By the Hanson-Wright inequality (e.g. Vershynin (2018, Theorem 6.2.1)), there exists a universal constant $c > 0$ such that for any $\epsilon \in (0,1)$,

$$\Pr\left(\left|\frac{\boldsymbol{Z}^\top \boldsymbol{P}_{\boldsymbol{u}^\perp}\boldsymbol{Z}}{p-1} - 1\right| > \epsilon\right) \leq 2\exp\left\{-c\epsilon^2(p-1)\right\}\,.$$

Now since for $\delta > 0$, $|z-1| \geq \delta$ implies that $|z^2 - 1| \geq \max\{\delta, \delta^2\}$ (see Vershynin (2018, Equation 3.2)), one gets that

$$\Pr\left(\left|\frac{\|\boldsymbol{P}_{\boldsymbol{u}^\perp}\boldsymbol{Z}\|_2}{\sqrt{p-1}} - 1\right| > \epsilon\right) \leq 2\exp\left\{-c\epsilon^2(p-1)\right\}\,.$$

By the first Borel-Cantelli Lemma, it follows that $\lim_{p\to\infty} \|\boldsymbol{P}_{\boldsymbol{u}^\perp}\boldsymbol{Z}\|_2/(p-1) = 1$ almost surely and the claim follows. $\qquad\square$

References

Beck A (2017). *First-order methods in optimization.* SIAM. ISBN 978-1-61197-498-0. doi:10.1137/1. 9781611974997.

Candès EJ, Sur P (2020). "The phase transition for the existence of the maximum likelihood estimate in high-dimensional logistic regression." *The Annals of Statistics*, **48**(1), 27–42. doi:10.1214/18-AOS1789.

Diaconis P, Ylvisaker D (1979). "Conjugate priors for exponential families." *The Annals of Statistics*, **7**(2), 269–281. doi:10.1214/aos/1176344611.

Donoho D, Montanari A (2016). "High dimensional robust M-estimation: Asymptotic variance via approximate message passing." *Probability Theory and Related Fields*, **166**, 935–969. doi:10.1007/ s00440-015-0675-z.

Feng OY, Venkataramanan R, Rush C, Samworth RJ (2022). "A unifying tutorial on approximate message passing." *Foundations and Trends in Machine Learning*, **15**(4), 335–536. doi:10.1561/2200000092.

Javanmard A, Montanari A (2013). "State evolution for general approximate message passing algorithms, with applications to spatial coupling." *Information and Inference: A Journal of the IMA*, **2**(2), 115–144. doi:10.1093/imaiai/iat004.

Magnus JR, Neudecker H (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley & Sons. ISBN 9781119541202. doi:10.1002/9781119541219.

Rudelson M, Vershynin R (2010). *Non-asymptotic theory of random matrices: Extreme singular values*, pp. 1576–1602. doi:10.1142/9789814324359_0111.

Rudin W (1976). *Principles of Mathematical Analysis.* McGraw-hill New York, 3rd edition. ISBN 978-0070542358.

Shiryaev AN (2016). *Probability-1*, volume 95. Springer. ISBN 978-1-4939-7905-9. doi:10.1007/ 978-0-387-72206-1.

Stein CM (1981). "Estimation of the mean of a multivariate normal distribution." *The Annals of Statistics*, **9**(6), 1135–1151. doi:10.1214/aos/1176345632.

Sur P (2019). *A Modern Maximum Likelihood Theory for High-dimensional Logistic Regression*. Ph.D. thesis, Stanford University. URL `http://purl.stanford.edu/jw604jq1260`.

Sur P, Candès EJ (2019). "A modern maximum-likelihood theory for high-dimensional logistic regression." *Proceedings of the National Academy of Sciences*, **116**(29), 14516–14525. doi:10.1073/pnas.1810420116.

Sur P, Chen Y, Candès EJ (2019). "The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square." *Probability theory and related fields*, **175**, 487–558. doi:10.1007/s00440-018-00896-9.

Vershynin R (2018). *High-dimensional Probability: An Introduction with Applications in Data Science*, volume 47. Cambridge university press.

Zehna PW (1966). "Invariance of maximum likelihood estimators." *The Annals of Mathematical Statistics*, **37**(3), 744–744. doi:10.1214/aoms/1177699475.

Zhao Q, Sur P, Candes EJ (2022). "The asymptotic distribution of the MLE in high-dimensional logistic models: Arbitrary covariance." *Bernoulli*, **28**(3), 1835–1861. doi:10.3150/21-BEJ1401.