RESTRICTED EIGENVALUES AND MAX-NORM BOUNDS FOR NORMAL-SCORE CORRELATION MATRICES

Anonymous authors
Paper under double-blind review

ABSTRACT

In a p-dimensional Gaussian copula model, the latent correlation matrix Σ determines the dependence structure while the marginal distributions remain unspecified. The normal-score, or van der Waerden, correlation matrix $\widehat{\Sigma}^{\mathrm{NS}}$ is the sample correlation matrix of the standard normal quantiles of the scaled ranks of n independent observations; it is asymptotically efficient in fixed dimension. A previous high-dimensional analysis of this estimator gives an entrywise expansion whose remainder bound constrains p relative to n. We show that, with high probability, $\widehat{\Sigma}^{\mathrm{NS}}$ differs from the oracle sample correlation matrix of the latent Gaussian data by at most a constant multiple of $(\log p + \log\log n)/n$ in max norm. When $\log p = o(n)$ this difference is negligible relative to the max-norm scale $\sqrt{\log p/n}$, so the entrywise and restricted-eigenvalue guarantees of the Gaussian oracle carry over to $\widehat{\Sigma}^{\mathrm{NS}}$. In particular, $\widehat{\Sigma}^{\mathrm{NS}}$ satisfies a restricted-eigenvalue condition whenever Σ does, once $n \gtrsim s\log p$, the Gaussian-design sample-size order; the generic conversion of a max-norm bound relative to Σ would instead require $n \gtrsim s^2 \log p$. Moreover, $\|\widehat{\Sigma}^{\mathrm{NS}} - \Sigma\|_{\max} \lesssim \sqrt{\log p/n}$ with no restriction on p relative to n, and this rate is sharp. For sparse targets of bounded ℓ_1 norm, normal-score procedures for Gaussian copula regression and inverse correlation estimation therefore attain the usual $\sqrt{s\log p/n}$ rate at Gaussian-design sample sizes, without the earlier dimensional restrictions. The proof couples each normalized score column with the latent Gaussian column. The column error is a one-dimensional matching error; its Euclidean norm is a Lipschitz function of the latent column with mean square of order $\log\log n$. Gaussian concentration therefore controls all columns at once.

1 INTRODUCTION

Let $X = (X_1, \ldots, X_p)^\top$ be a p-dimensional random vector following a Gaussian copula model: there are unknown strictly increasing marginal transformations $f_1, \ldots, f_p$ such that

$$Z = f(X) := (f_1(X_1), \ldots, f_p(X_p))^\top \sim N_p(0, \Sigma), \qquad \mathrm{diag}(\Sigma) = \mathbf{1}. \tag{1}$$

Here Σ is the correlation matrix of the latent Gaussian vector Z, and the marginal distributions of X are unspecified; our goal is to estimate Σ from n independent observations of X. The diagonal normalization identifies the representation: $f_j = \Phi^{-1} \circ F_j$, where F_j is the distribution function of X_j and Φ is the standard normal distribution function. This is the nonparanormal model of Liu et al. (2009).

Because the transformations are unknown, estimation of Σ starts from ranks, and two families of rank-based estimators are in use. The first converts Kendall's tau and Spearman's rho into estimates of Σ_{jk} through the sine transforms valid in the Gaussian copula model (Liu et al., 2012). Their high-dimensional theory is well developed. Liu et al. (2012) prove entrywise concentration bounds. Mitra & Zhang (2014) obtain spectral-norm bounds in the Gaussian copula model, including bounds over sparse submatrices. Barber & Kolar (2018) obtain sparse spectral bounds for the sine-transformed Kendall estimator in the transelliptical model, with dependence on the condition number of the latent correlation matrix. In high dimension, sparse regression coefficients and sparse inverse correlation matrices are estimated by Lasso-type ℓ_1-regularized quadratic programs whose Hessian is the estimated latent correlation matrix. The sine transforms need not produce a positive semidefinite matrix, so such programs may require a projection step (Liu et al., 2012; Zhao et al., 2012; Xue & Zou, 2012; Cai & Zhang, 2018; Zhao & Genest, 2019).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

The second family is the normal-score rank correlation estimator, also known as the van der Waerden correlation matrix (Zhao & Genest, 2019). Let R_{ij} be the rank of X_{ij} among $X_{1j}, \ldots, X_{nj}$, and define

$$q_r = \Phi^{-1}\left(\frac{r}{n+1}\right), \quad r = 1, \ldots, n, \qquad \gamma_n = \frac{1}{n}\sum_{r=1}^{n} q_r^2, \tag{2}$$

together with

$$\widehat{Z}_{ij} = q_{R_{ij}}, \qquad \widehat{\Sigma}^{\mathrm{NS}} = \frac{1}{n\gamma_n}\widehat{Z}^\top \widehat{Z}. \tag{3}$$

The normalization by γ_n gives (3) unit diagonal and the Gram form makes it positive semidefinite, so no projection is needed. For fixed dimension and positive definite Σ, the estimator is asymptotically semiparametrically efficient in the unrestricted Gaussian copula model, with normal margins least favorable (Klaassen & Wellner, 1997; Segers et al., 2014).

The high-dimensional theory of normal scores is less complete, for two reasons. The Gaussian quantile function is unbounded at the endpoints, so the extreme scores grow with n, and all scores in one column are dependent because every rank uses the whole column. The arguments for the sine-transformed estimators do not transfer to normal scores. Early nonparanormal work avoided the endpoints through Winsorization at level $\delta_n \asymp n^{-1/4}(\log n)^{-1/2}$ (Liu et al., 2009); the results below concern the untruncated score (2), for which $\delta_n = 1/(n+1)$. Zhao & Genest (2019, Appendix F) established a uniform asymptotic linear expansion for the untruncated estimator under $\log p = o(n)$ and $\log n = O(p)$. Their stated remainder bound is negligible relative to $\sqrt{\log p/n}$ when $\log p = o(\sqrt{n})$ but not throughout $\log p = o(n)$ (Appendix B).

The Lasso-type programs above, with either estimate of Σ as Hessian, need two properties of that matrix: entrywise bounds control the loss gradient at the target, and restricted eigenvalues its curvature. Restricted eigenvalues can be extracted from an entrywise bound through the generic inequality $|v^\top(\widehat{\Sigma}^{\mathrm{NS}} - \Sigma)v| \le \left\|\widehat{\Sigma}^{\mathrm{NS}} - \Sigma\right\|_{\max} \|v\|_1^2$, which on an s-sparse cone multiplies the entrywise error by a factor of order s (Bickel et al., 2009). For the deviation from Σ, of order $\sqrt{\log p/n}$, a fixed cone constant and population restricted eigenvalues bounded away from zero then require a sample size of order $s^2 \log p$.

Our central result, Theorem 2.1, is an oracle approximation. Let $\widehat{R}^G$ be the Pearson correlation matrix of the centered latent Gaussian sample, which cannot be computed without knowing the marginal distributions. Then, with high probability,

$$\left\|\widehat{\Sigma}^{\mathrm{NS}} - \widehat{R}^G\right\|_{\max} \lesssim \frac{\log p + \log\log n}{n}.$$

The right side is of order $(\log p + \log\log n)/n$ rather than $\sqrt{\log p/n}$, so the same generic inequality transfers restricted eigenvalues from the oracle to $\widehat{\Sigma}^{\mathrm{NS}}$ at a sample size of order $s \log p$.

Combined with a Gaussian-design bound for $\widehat{R}^G$, Theorem 2.1 compares the design norms $\sqrt{v^\top \widehat{\Sigma}^{\mathrm{NS}} v}$ and $\sqrt{v^\top \Sigma v}$ uniformly over all v (Corollary 2.3):

$$\left|\sqrt{v^\top \widehat{\Sigma}^{\mathrm{NS}} v} - \sqrt{v^\top \Sigma v}\right| \lesssim \sqrt{\frac{\log p + \log\log n}{n}}\ \|v\|_1$$

with high probability. On an s-sparse cone the sample restricted eigenvalue therefore lies within a constant multiple of $\sqrt{s(\log p + \log\log n)/n}$ of the population one. When the latter is bounded away from zero, the two agree up to a constant factor as close to one as desired (Corollary 2.4) once $n \gtrsim s \log p$, the sample-size order of Gaussian designs (Raskutti et al., 2010). The additive $\log\log n$ term is absorbed into the sample-size constant, which may depend on the prescribed relative accuracy (Corollary 2.4).

The second consequence, Corollary 2.7, is the max-norm bound

$$\left\|\widehat{\Sigma}^{\mathrm{NS}} - \Sigma\right\|_{\max} \lesssim \sqrt{\frac{\log p}{n}} \tag{4}$$

with high probability and with no restriction on p relative to n. It combines Theorem 2.1 with an entrywise bound for the oracle; the $\log\log n$ term, being of lower order, does not appear in the rate. The rate in (4) is sharp over the class of correlation matrices: when $\Sigma = I_p$ and $n, p \to \infty$ with $\log p = o(n)$, $\sqrt{n/\log p}\ \left\|\widehat{\Sigma}^{\mathrm{NS}} - \Sigma\right\|_{\max}$ converges in probability to 2 (Proposition 2.8).

All results rest on one idea. Rather than linearizing the rank transform entry by entry, we view the normalized score matrix as a perturbation of the latent Gaussian data matrix. Each normalized score column is the latent column with its entries replaced by the closest rearrangement of a fixed, normalized grid of Gaussian quantiles, so the column error is a one-dimensional matching error. Its Euclidean norm is a 1-Lipschitz function of the latent column, and its mean square is only of order $\log\log n$, a vanishing fraction of the squared column norm n. Gaussian concentration therefore controls all p column-error norms at once. The column errors enter the Gram matrix through a cross term and a quadratic term. Expanding $\widehat{\Sigma}^{\mathrm{NS}} - \Sigma$ gives a leading term plus a remainder of order $(\log p + \log\log n)/n$ in max norm (Section 3.2). When Σ is positive definite, the leading term is the average of the efficient influence functions at the latent observations. The oracle admits the same expansion with a smaller remainder, and the two leading terms cancel, which gives Theorem 2.1.

These results remove the dimensional restriction that the remainder bound of Zhao & Genest (2019) imposed on two applications, Gaussian copula regression and sparse inverse correlation estimation. Their ℓ_1-penalized programs use a normal-score matrix as Hessian, positive semidefinite without projection. Both attain the usual $\sqrt{s\log p/n}$ rate for targets of bounded ℓ_1 norm under the sample-size condition $n \gtrsim s\log p$ and a population restricted-eigenvalue bound (Section 4).

Roadmap. Section 2 states the main results. Section 3 proves them: Section 3.1 constructs the coupling, Section 3.2 proves Theorem 2.1, and Sections 3.3 and 3.4 derive the restricted-eigenvalue and max-norm corollaries from it and from two deviation bounds for the Gaussian oracle. Proofs of the lemmas are in Appendices A to C. Section 4 gives the applications, whose proofs are in Appendix D; Section 5 discusses the approach, its implications and an open question.

Notation. Throughout, n is the sample size, p the dimension and s the sparsity level, and $[p] = \{1,\dots,p\}$. Tail bounds are stated as holding with probability at least $1 - Ce^{-t}$ for a tail parameter $t \geq 0$, where c, C denote positive universal constants whose value may change between statements but never depends on n, p, s, t or Σ. Replacing t by $u + \log C$ turns $1 - Ce^{-t}$ into $1 - e^{-u}$ at the cost of enlarging C. As usual, $a \lesssim b$ means $a \leq Cb$, $a \gtrsim b$ means $b \lesssim a$, and $a \asymp b$ means both. For a matrix H, $\|H\|_{\max} = \max_{j,k}|H_{jk}|$, $\mathrm{diag}(H)$ is the vector of its diagonal entries, and $\mathrm{Diag}(a_1,\dots,a_p)$ is the diagonal matrix with the indicated entries. The vector $\mathbf{1}$ has all entries equal to one, and $\|v\|_0$ is the number of nonzero coordinates of v.

2 Main results

Since the marginal transformations in (1) are strictly increasing, the observed and latent samples have identical within-coordinate ranks. We may therefore analyze (3) as a function of the latent sample $Z_1,\dots,Z_n$ and compare it with the sample correlation matrix of that sample. Section 2.1 states the main result: with high probability, the two matrices differ by at most a constant multiple of $(\log p + \log\log n)/n$ in max norm. Sections 2.2 and 2.3 derive from it restricted-eigenvalue comparisons at the Gaussian-design sample size $n \gtrsim s\log p$, and the max-norm rate $\sqrt{\log p/n}$. Throughout, $X_1,\dots,X_n$ are iid from the continuous Gaussian copula model (1), with $n \geq 3$ and $p \geq 2$ so that $\log\log n$ and $\log p$ are positive.

2.1 Oracle approximation

Let $P_n = I_n - n^{-1}\mathbf{1}\mathbf{1}^\top$, let $\widehat{\Sigma}^G = n^{-1}Z^\top P_n Z$ be the centered latent Gaussian sample covariance matrix, and define the corresponding correlation matrix $\widehat{R}^G$, the sample correlation matrix one would compute if the latent Gaussian data were observed, by $\widehat{R}^G_{jk} = \widehat{\Sigma}^G_{jk}/(\widehat{\Sigma}^G_{jj}\widehat{\Sigma}^G_{kk})^{1/2}$ for $j,k \in [p]$.

Theorem 2.1 (Second-order oracle approximation)**.** *For any $t \geq 0$, with probability at least $1 - Ce^{-t}$, we have*

$$\left\|\widehat{\Sigma}^{\mathrm{NS}} - \widehat{R}^G\right\|_{\max} \leq C\frac{\log p + t + \log\log n}{n}.$$

When $\log p + t = o(n)$ the right side is negligible relative to the max-norm scale $\sqrt{(\log p + t)/n}$, so the two estimators share the leading term of their max-norm expansions. For positive definite Σ this common term is the average of the efficient influence functions. The $\log\log n$ term is the

cost of the coupling (Appendix A); it appears in the remainder bounds, not in the leading max-norm rate. The oracle approximation, combined with Gaussian-design bounds for the oracle, yields the restricted-eigenvalue and max-norm guarantees below. Theorem 2.1 is proved in Section 3.2.

2.2 Restricted eigenvalues

Definition 2.2 (Cone and restricted eigenvalues)**.** For $S \subset [p]$ and $c_0 \geq 1$, let

$$\mathcal{C}(S, c_0) = \{v \in \mathbb{R}^p : \|v_{S^c}\|_1 \leq c_0 \|v_S\|_1\},$$

and for a symmetric positive semidefinite matrix H define

$$\kappa_H(s, c_0) = \min_{1 \leq |S| \leq s} \inf_{v \in \mathcal{C}(S,c_0) \setminus \{0\}} \frac{\sqrt{v^\top H v}}{\|v\|_2}.$$

This is the normalization of Raskutti et al. (2010); Cai & Zhang (2018) use the squared constant.

Corollary 2.3 (Two-sided comparison of the normal-score and population quadratic forms)**.** *For any $t \geq 0$, with probability at least $1 - Ce^{-t}$, we have*

$$\left|\sqrt{v^\top \widehat{\Sigma}^{\mathrm{NS}} v} - \sqrt{v^\top \Sigma v}\right| \leq C\sqrt{\frac{\log p + t + \log\log n}{n}}\ \|v\|_1 \qquad \text{simultaneously for all } v \in \mathbb{R}^p. \tag{5}$$

On the same event, for any $\eta \in (0, 1)$ and all $v \in \mathbb{R}^p$,

$$\left|v^\top \widehat{\Sigma}^{\mathrm{NS}} v - v^\top \Sigma v\right| \leq \eta\, v^\top \Sigma v + \frac{2C^2}{\eta}\, \frac{\log p + t + \log\log n}{n}\, \|v\|_1^2. \tag{6}$$

Corollary 2.3 states that for any direction v the normal-score design norm $\sqrt{v^\top \widehat{\Sigma}^{\mathrm{NS}} v}$ agrees with $\sqrt{v^\top \Sigma v}$ up to a multiple of $\|v\|_1 \sqrt{(\log p + t + \log\log n)/n}$. For the latent Gaussian design itself the same comparison holds with $\log p + t$ alone. The ℓ_1 norm on the right side is what a sparse cone can absorb, which is how the next corollary transfers restricted eigenvalues.

Corollary 2.4 (Restricted-eigenvalue approximation)**.** *On the event of (5), for any $1 \leq s \leq p$ and $c_0 \geq 1$,*

$$\left|\kappa_{\widehat{\Sigma}^{\mathrm{NS}}}(s, c_0) - \kappa_\Sigma(s, c_0)\right| \leq C(1 + c_0)\sqrt{\frac{s(\log p + t + \log\log n)}{n}}. \tag{7}$$

Consequently, for any $\kappa_0 > 0$, $c_0 \geq 1$ and $\eta \in (0, 1)$ there is a constant A depending only on κ_0, c_0 and η such that

$$\kappa_\Sigma(s, c_0) \geq \kappa_0, \qquad n \geq A\, s \log p \tag{8}$$

imply that, with probability at least $1 - C/p$, $(1 - \eta)\, \kappa_\Sigma(s, c_0) \leq \kappa_{\widehat{\Sigma}^{\mathrm{NS}}}(s, c_0) \leq (1 + \eta)\, \kappa_\Sigma(s, c_0)$.

Thus the normal-score matrix satisfies a restricted-eigenvalue condition whenever Σ does, at the Gaussian-design sample size, so the standard restricted-eigenvalue analysis of ℓ_1 programs applies (Section 4).

Theorem 2.1 alone also bounds the change in restricted eigenvalues from the oracle to $\widehat{\Sigma}^{\mathrm{NS}}$.

Corollary 2.5 (Oracle restricted eigenvalues)**.** *For any $t \geq 0$, with probability at least $1 - Ce^{-t}$, we have, for all $1 \leq s \leq p$ and $c_0 \geq 1$,*

$$\left|\kappa_{\widehat{\Sigma}^{\mathrm{NS}}}(s, c_0)^2 - \kappa_{\widehat{R}^G}(s, c_0)^2\right| \leq C(1 + c_0)^2\, s\, \frac{\log p + t + \log\log n}{n}.$$

This bound compares the squared restricted eigenvalues of the two sample matrices directly. A comparison with the population value κ_Σ, as in Corollary 2.4, additionally requires a deviation bound for the Gaussian oracle.

Corollary 2.6 (Two-sided sparse control)**.** *On the event of (5),*

$$\sup_{\|v\|_2 = 1,\ \|v\|_0 \leq s} \left|v^\top (\widehat{\Sigma}^{\mathrm{NS}} - \Sigma) v\right| \leq 2\sqrt{\Lambda_s}\, r\sqrt{s} + r^2 s,$$

where $r = C\sqrt{(\log p + t + \log\log n)/n}$ and $\Lambda_s = \max_{|S| \leq s} \lambda_{\max}(\Sigma_{SS})$.

When Λ_s is bounded and $s(\log p + \log\log n) \lesssim n$, this bound is of order $\sqrt{s(\log p + \log\log n)/n}$, whereas the latent Gaussian sample covariance has a bound of order $\sqrt{s \log(ep/s)/n}$ (Mitra & Zhang, 2014, eq. (1.2)). The two orders agree whenever $\log p + \log\log n \lesssim \log(ep/s)$. Corollaries 2.3 and 2.5 are proved in Section 3.3. The proofs of Corollaries 2.4 and 2.6 are in Appendix C.

2.3 MAX-NORM RATE

Corollary 2.7 (Max-norm rate). *For any $t \geq 0$, with probability at least $1 - Ce^{-t}$, we have*

$$\left\|\widehat{\Sigma}^{\mathrm{NS}} - \Sigma\right\|_{\max} \leq C\left\{\sqrt{\frac{\log p + t}{n}} + \frac{\log p + t + \log\log n}{n}\right\}. \tag{9}$$

On the same event, (9) implies (see Section 3.4)

$$\left\|\widehat{\Sigma}^{\mathrm{NS}} - \Sigma\right\|_{\max} \leq C\sqrt{\frac{\log p + t}{n}}. \tag{10}$$

Corollary 2.7 gives the canonical max-norm rate $\sqrt{(\log p + t)/n}$. The additional term in (9) is negligible relative to this scale when $\log p + t = o(n)$. The corollary holds for every correlation matrix and for all p and n; bound (10) carries information throughout the range $\log p + t \lesssim n$, which permits p exponential in n.

Proposition 2.8 (Sharpness of the max-norm rate). *Suppose that $\Sigma = I_p$ and $n, p \to \infty$ with $\log p = o(n)$. Then*

$$\sqrt{\frac{n}{\log p}}\,\left\|\widehat{\Sigma}^{\mathrm{NS}} - I_p\right\|_{\max} \longrightarrow 2 \qquad \textit{in probability.}$$

In particular the order in (10) is exact.

Although Corollary 2.7 gives the sharp entrywise rate, combining it with the generic inequality $|v^\top(\widehat{\Sigma}^{\mathrm{NS}} - \Sigma)v| \leq \left\|\widehat{\Sigma}^{\mathrm{NS}} - \Sigma\right\|_{\max} \|v\|_1^2$ yields only $n \gtrsim s^2 \log p/\kappa_\Sigma^4$, because the deviation from Σ is of order $\sqrt{\log p/n}$. The oracle deviation in Theorem 2.1 is of order $(\log p + t + \log\log n)/n$, and the same inequality transfers restricted eigenvalues from $\widehat{R}^G$ to $\widehat{\Sigma}^{\mathrm{NS}}$ at the order $s \log p$ (Corollary 2.5).

Corollary 2.7 is proved in Section 3.4. The proof of Proposition 2.8 is in Appendix C.

3 PROOFS OF THE MAIN RESULTS

Section 3.1 constructs the coupling; Section 3.2 proves Theorem 2.1; Sections 3.3 and 3.4 derive the corollaries. Auxiliary proofs are in Appendices A to C. At a fixed tail level t, intersections of a fixed number of the stated high-probability events have probability at least $1 - Ce^{-t}$, after increasing C.

3.1 NORMAL-SCORE COUPLING

Ranks retain the ordering within each coordinate, but not the observed values. Normal scoring replaces them by fixed Gaussian-quantile scores in the same order. The ordered latent observations fluctuate around Gaussian quantiles, whereas the normalized score grid is deterministic. We compare the two through the Euclidean norm of the error in each score column, rather than through a uniform coordinatewise bound. This norm is the distance from the latent column to a fixed set, hence a 1-Lipschitz function of the latent column. A one-dimensional quantile comparison bounds its expected square by a multiple of $\log\log n$. Both facts, proved in Appendix A, give the uniform column bound below.

Define

$$\widetilde{Z} = \gamma_n^{-1/2}\widehat{Z}, \qquad E = \widetilde{Z} - Z,$$

so that $\widehat{\Sigma}^{\mathrm{NS}} = n^{-1}\widetilde{Z}^\top\widetilde{Z}$ and $\widetilde{Z} = Z + E$; E is the coupling error. For $z \in \mathbb{R}^n$ with distinct coordinates, let $\mathcal{N}_n(z)$ assign $q_r/\sqrt{\gamma_n}$ to the coordinate of rank r. Since $X_{\cdot j}$ and $Z_{\cdot j}$ have the same ranks, $\widetilde{Z}_{\cdot j} = \mathcal{N}_n(Z_{\cdot j})$ almost surely. Two consequences of the construction are used in Section 3.2: the coupling error $E_{\cdot j} = \mathcal{N}_n(Z_{\cdot j}) - Z_{\cdot j}$ is a function of $Z_{\cdot j}$ alone, and every column of $\widetilde{Z}$ has squared Euclidean norm $\gamma_n^{-1}\sum_r q_r^2 = n$ exactly.

Lemma 3.1 (Uniform columnwise coupling). *For any $t \geq 0$, with probability at least $1 - e^{-t}$, we have*

$$\max_{j \leq p} \|E_{\cdot j}\|_2 \leq C\{\sqrt{\log\log n} + \sqrt{\log p + t}\}. \tag{11}$$

The $\log\log n$ term comes from the mean-square bound of Appendix A, which also shows that this order is exact.

3.2 Proof of Theorem 2.1

We expand both matrices around Σ to first order: the expansions share their leading term, the coupling controls the remainder of the normal-score expansion, and Gaussian moment bounds control that of the oracle. With $\widetilde{Z} = Z + E$ as in Section 3.1, define $\widehat{\Sigma}_0^G = n^{-1}Z^\top Z$, $A = n^{-1}Z^\top E$ and $B = n^{-1}E^\top E$, so that $\widehat{\Sigma}^{\mathrm{NS}} = \widehat{\Sigma}_0^G + A + A^\top + B$. Here A is the cross term between the latent design and the coupling error, and B the quadratic coupling-error term.

Both expansions have leading term $n^{-1}\sum_{i=1}^n \Psi_\Sigma(Z_i)$, where for $z \in \mathbb{R}^p$

$$\Psi_\Sigma(z) = zz^\top - \frac{1}{2}\left\{\mathrm{Diag}(z_1^2,\dots,z_p^2)\Sigma + \Sigma\,\mathrm{Diag}(z_1^2,\dots,z_p^2)\right\}, \tag{12}$$

with (j,k) entry $\psi_{jk}(z) = z_j z_k - \frac{1}{2}\Sigma_{jk}(z_j^2 + z_k^2)$. This is the first-order effect of one latent observation z on a correlation matrix: $zz^\top - \Sigma$ corrected for the normalization of the diagonal. Its diagonal entries vanish and its mean under $N_p(0,\Sigma)$ is zero. When Σ is positive definite, its (j,k) entry is the efficient influence function for Σ_{jk} in the Gaussian copula model (Klaassen & Wellner, 1997; Segers et al., 2014); the definition and the bounds below apply to all correlation matrices Σ.

Conditioning $Z_{\cdot k}$ on $Z_{\cdot j}$ splits A_{kj} into $\Sigma_{kj}A_{jj}$ and a residual that is centered Gaussian given $Z_{\cdot j}$, because $E_{\cdot j}$ depends only on $Z_{\cdot j}$. Let Δ_0 and Δ_B be the diagonal matrices with entries $(\widehat{\Sigma}_0^G)_{jj} - 1$ and B_{jj}. The fixed score-column norms give $A_{jj} = -\frac{1}{2}(\Delta_0)_{jj} - \frac{1}{2}B_{jj}$. In $A + A^\top$, the Δ_0 contributions combine with $\widehat{\Sigma}_0^G - \Sigma$ to give $n^{-1}\sum_i \Psi_\Sigma(Z_i)$; the Δ_B contributions are quadratic in E.

Proposition 3.2 (Conditional decomposition of the cross term)**.** *For $j,k \in [p]$, define $\Xi_{kj} = n^{-1}(Z_{\cdot k} - \Sigma_{kj}Z_{\cdot j})^\top E_{\cdot j}$. Then $\Xi_{jj} = 0$ and*

$$\widehat{\Sigma}^{\mathrm{NS}} - \Sigma = \frac{1}{n}\sum_{i=1}^n \Psi_\Sigma(Z_i) + R_n, \qquad R_n = \Xi + \Xi^\top + B - \frac{1}{2}(\Delta_B\Sigma + \Sigma\Delta_B). \tag{13}$$

Moreover, conditionally on $Z_{\cdot j}$,

$$\Xi_{kj} \mid Z_{\cdot j} \sim N\left(0, \frac{1-\Sigma_{jk}^2}{n^2}\,\|E_{\cdot j}\|_2^2\right). \tag{14}$$

Given $Z_{\cdot j}$, each Ξ_{kj} is Gaussian with standard deviation at most $n^{-1}\|E_{\cdot j}\|_2$ by (14). Gaussian tails at level $t + 2\log p$, a union bound over the p^2 entries and Lemma 3.1 give the following bound; the cross product of $\sqrt{\log p + t}$ and $\sqrt{\log\log n}$ is absorbed by $2ab \le a^2 + b^2$.

Lemma 3.3 (Residual cross terms)**.** *For any $t \ge 0$, with probability at least $1 - Ce^{-t}$, we have*

$$\|\Xi\|_{\max} \le C\frac{\log p + t + \log\log n}{n}.$$

Proposition 3.4 (Nonasymptotic efficient expansion)**.** *For any $t \ge 0$, with probability at least $1 - Ce^{-t}$, we have*

$$\left\|\widehat{\Sigma}^{\mathrm{NS}} - \Sigma - \frac{1}{n}\sum_{i=1}^n \Psi_\Sigma(Z_i)\right\|_{\max} \le C\frac{\log p + t + \log\log n}{n}. \tag{15}$$

By Proposition 3.2, it suffices to bound R_n. On the event of Lemma 3.1, since $B \succeq 0$,

$$\|B\|_{\max} \le \max_j B_{jj} = \frac{1}{n}\max_j \|E_{\cdot j}\|_2^2 \le C\frac{\log p + t + \log\log n}{n}.$$

Since Δ_B is diagonal and $|\Sigma_{jk}| \le 1$, $\|\Delta_B\Sigma + \Sigma\Delta_B\|_{\max} \le 2\|B\|_{\max}$. Thus (13) gives $\|R_n\|_{\max} \le 2\|\Xi\|_{\max} + 2\|B\|_{\max}$. Intersecting with the event of Lemma 3.3 gives (15) and proves Proposition 3.4.

Expanding the oracle's sample-variance normalization to second order isolates the same leading term. The expansion requires the sample variances to lie within a constant factor of one, which holds on an event of probability at least $1 - Ce^{-t}$ once $\log p + t \le cn$; the fixed score-column norms spare $\widehat{\Sigma}^{\mathrm{NS}}$ this step. On the same event the sample means are at most $C\sqrt{(\log p + t)/n}$ in absolute value, so the centering correction, which consists of their products, contributes at most $C(\log p + t)/n$.

Lemma 3.5 (Oracle expansion). *There is a universal constant $c > 0$ such that, if $\log p + t \le cn$, then with probability at least $1 - Ce^{-t}$ we have*

$$\left\| \widehat{R}^G - \Sigma - \frac{1}{n}\sum_{i=1}^{n} \Psi_\Sigma(Z_i) \right\|_{\max} \le C\frac{\log p + t}{n}.$$

Let c be the constant of Lemma 3.5. If $\log p + t > cn$, both matrices are correlation matrices, so $\left\|\widehat{\Sigma}^{\mathrm{NS}} - \widehat{R}^G\right\|_{\max} \le 2 \le (2/c)(\log p + t)/n$. Assume $\log p + t \le cn$. Then

$$\widehat{\Sigma}^{\mathrm{NS}} - \widehat{R}^G = \left(\widehat{\Sigma}^{\mathrm{NS}} - \Sigma - \frac{1}{n}\sum_{i=1}^{n} \Psi_\Sigma(Z_i)\right) - \left(\widehat{R}^G - \Sigma - \frac{1}{n}\sum_{i=1}^{n} \Psi_\Sigma(Z_i)\right),$$

and on the intersection of the events of Proposition 3.4 and Lemma 3.5, the max norm of the right side is at most $C(\log p + t + \log\log n)/n$. This proves Theorem 2.1.

3.3 Proofs of the restricted-eigenvalue corollaries

Corollary 2.3 compares $\widehat{\Sigma}^{\mathrm{NS}}$ with Σ, and Theorem 2.1 compares it with $\widehat{R}^G$; the missing comparison is that of the oracle with Σ.

Lemma 3.6 (Design-norm comparison for the oracle). *For any $t \ge 0$, with probability at least $1 - Ce^{-t}$, we have*

$$\left|\sqrt{v^\top \widehat{R}^G v} - \sqrt{v^\top \Sigma v}\right| \le C\sqrt{\frac{\log p + t}{n}}\ \|v\|_1 \qquad \textit{simultaneously for all } v \in \mathbb{R}^p.$$

The matrix-deviation inequality of Liaw et al. (2017) gives this comparison for the uncentered Gram matrix $n^{-1}Z^\top Z$. Centering and column normalization each add at most $C\sqrt{(\log p + t)/n}\ \|v\|_1$.

For Corollary 2.3, fix $t \ge 0$. For every $v \in \mathbb{R}^p$, $|v^\top(\widehat{\Sigma}^{\mathrm{NS}} - \widehat{R}^G)v| \le \left\|\widehat{\Sigma}^{\mathrm{NS}} - \widehat{R}^G\right\|_{\max} \|v\|_1^2$, and for $x, y \ge 0$ we have $(\sqrt{x} - \sqrt{y})^2 \le |x - y|$. Both matrices are positive semidefinite, so on the event of Theorem 2.1

$$\left|\sqrt{v^\top \widehat{\Sigma}^{\mathrm{NS}} v} - \sqrt{v^\top \widehat{R}^G v}\right| \le C\sqrt{\frac{\log p + t + \log\log n}{n}}\ \|v\|_1 \qquad \text{for all } v \in \mathbb{R}^p. \tag{16}$$

On the intersection of this event with that of Lemma 3.6, the triangle inequality gives (5).

For (6), let $a = \sqrt{v^\top \Sigma v}$ and let $b = C\sqrt{(\log p + t + \log\log n)/n}\ \|v\|_1$, so that $(a - b)_+ \le \sqrt{v^\top \widehat{\Sigma}^{\mathrm{NS}} v} \le a + b$. For $\eta \in (0, 1)$, by $2ab \le \eta a^2 + \eta^{-1} b^2$, $(a+b)^2 \le (1+\eta)a^2 + (1+\eta^{-1})b^2$ and $(a-b)_+^2 \ge (1-\eta)a^2 - (\eta^{-1} - 1)b^2$ (for $a < b$ the right side is nonpositive), and $1 + \eta^{-1} \le 2\eta^{-1}$; together these give (6). This proves Corollary 2.3.

For Corollaries 2.4 and 2.5, let $v \in \mathcal{C}(S, c_0)$ be nonzero with $|S| \le s$, so that $\|v\|_1 \le (1+c_0)\sqrt{s}\ \|v\|_2$. On the event of (5), division by $\|v\|_2$ bounds the difference of the normal-score and population ratios by $C(1 + c_0)\sqrt{s(\log p + t + \log\log n)/n}$, the right side of (7). On the event of Theorem 2.1, the same cone inequality gives $|v^\top(\widehat{\Sigma}^{\mathrm{NS}} - \widehat{R}^G)v| / \|v\|_2^2 \le C(1+c_0)^2 s(\log p + t + \log\log n)/n$. Infima over nonzero cone vectors and minima over supports then give the bounds in Corollaries 2.4 and 2.5. For Corollary 2.6, use $\|v\|_1 \le \sqrt{s}$ and $v^\top \Sigma v \le \Lambda_s$ on sparse unit vectors. The last statement of Corollary 2.4 absorbs $\log\log n$ into the sample-size constant (Appendix C).

3.4 Proof of the max-norm corollary

Corollary 2.7 combines Theorem 2.1 with an entrywise bound for the oracle. Testing $v = e_j + e_k$ in Lemma 3.6 gives that bound, since both matrices have unit diagonal.

Lemma 3.7 (Max-norm bound for the oracle). *For any $t \ge 0$, with probability at least $1 - Ce^{-t}$, we have*

$$\left\| \widehat{R}^G - \Sigma \right\|_{\max} \le C\sqrt{\frac{\log p + t}{n}}.$$

On the intersection of the events of Theorem 2.1 and Lemma 3.7,

$$\left\|\widehat{\Sigma}^{\mathrm{NS}} - \Sigma\right\|_{\max} \le \left\|\widehat{\Sigma}^{\mathrm{NS}} - \widehat{R}^G\right\|_{\max} + \left\|\widehat{R}^G - \Sigma\right\|_{\max}$$

$$\leq C\left\{\frac{\log p + t + \log\log n}{n} + \sqrt{\frac{\log p + t}{n}}\right\},$$

which is (9). For (10), since $(\log p + t)/n \leq \sqrt{(\log p + t)/n}$ whenever $\log p + t \leq n$, while $\left\|\widehat{\Sigma}^{\mathrm{NS}} - \Sigma\right\|_{\max} \leq 2 \leq 2\sqrt{(\log p + t)/n}$ whenever $\log p + t > n$, and $\log\log n \leq 2\sqrt{n \log 2} \leq 2\sqrt{n(\log p + t)}$, the bound (9) implies (10) on the same event. This proves Corollary 2.7. Finally, Proposition 2.8 transfers the limit of Cai & Jiang (2012) for $\widehat{R}^G$ to $\widehat{\Sigma}^{\mathrm{NS}}$ through Theorem 2.1 (Appendix C).

4 Applications

We consider estimation of sparse regression coefficients and of sparse inverse correlation matrices in Gaussian copula models. Both procedures minimize ℓ_1-penalized quadratic losses whose Hessian is the normal-score matrix or a principal submatrix of it. As in the analysis of regularized M-estimators by Negahban et al. (2012), the error bounds need two inputs. The regularization parameter must dominate the ℓ_∞ norm of the loss gradient at the target; Corollary 2.7 supplies this bound. The Hessian must satisfy restricted strong convexity, here a restricted-eigenvalue bound on a sparse cone; Corollary 2.4 supplies it. The Hessian is positive semidefinite, so both programs are convex without a projection step, and the deterministic step from the two inputs to the error rates is the standard one, recorded in Appendix D. Throughout this section, A denotes the constant of Corollary 2.4 for $c_0 = 3$ and $\eta = 1/2$; it depends only on the lower bound κ_0 on the population restricted eigenvalue.

4.1 Gaussian copula regression

Let $(Y, X^\top)^\top$ follow (1) in dimension $p+1$, with $p \geq 2$ covariates; write Σ_{XX} for the covariate block of Σ, Σ_{XY} for the covariate–response column, partition $\widehat{\Sigma}^{\mathrm{NS}}$ accordingly, and let $\beta^\star$ be a solution of the normal equations $\Sigma_{XX}\beta^\star = \Sigma_{XY}$ for the regression of the latent response on the latent covariates. The normal-score estimator solves

$$\widehat{\beta} \in \arg\min_{\beta \in \mathbb{R}^p}\left\{\frac{1}{2}\beta^\top \widehat{\Sigma}^{\mathrm{NS}}_{XX}\beta - (\widehat{\Sigma}^{\mathrm{NS}}_{XY})^\top\beta + \lambda\|\beta\|_1\right\},$$

the normal-score version of the Kendall-based program of Cai & Zhang (2018); the objective is a penalized least-squares criterion in the score columns, so a minimizer exists. The gradient of the quadratic loss at $\beta^\star$ is $\widehat{\Sigma}^{\mathrm{NS}}_{XX}\beta^\star - \widehat{\Sigma}^{\mathrm{NS}}_{XY}$; since $\Sigma_{XX}\beta^\star = \Sigma_{XY}$,

$$\left\|\widehat{\Sigma}^{\mathrm{NS}}_{XX}\beta^\star - \widehat{\Sigma}^{\mathrm{NS}}_{XY}\right\|_\infty \leq (1 + \|\beta^\star\|_1)\left\|\widehat{\Sigma}^{\mathrm{NS}} - \Sigma\right\|_{\max}. \tag{17}$$

Corollary 4.1 (Gaussian copula regression)**.** *Suppose that* $1 \leq s \leq p$, $\|\beta^\star\|_0 \leq s$, $\|\beta^\star\|_1 \leq B$, $\kappa_{\Sigma_{XX}}(s,3) \geq \kappa_0 > 0$ *and* $n \geq A\, s\log p$. *Then the choice* $\lambda = C_\lambda(1+B)\sqrt{\log p/n}$, *with* C_λ *a sufficiently large universal constant, gives, with probability at least* $1 - C/p$,

$$\left\|\widehat{\beta} - \beta^\star\right\|_2 \leq \frac{C(1+B)}{\kappa_0^2}\sqrt{\frac{s\log p}{n}}, \qquad \left\|\widehat{\beta} - \beta^\star\right\|_1 \leq \frac{C(1+B)s}{\kappa_0^2}\sqrt{\frac{\log p}{n}}. \tag{18}$$

The proof follows the template: Corollary 2.7 with $t = \log p$ bounds the right side of (17) by a constant times $(1+B)\sqrt{\log p/n}$, which fixes λ; Corollary 2.4 with $\eta = 1/2$ gives $\kappa_{\widehat{\Sigma}^{\mathrm{NS}}_{XX}}(s,3) \geq \kappa_0/2$ when $n \geq A\, s\log p$; and the deterministic bound of Appendix D then gives (18).

The assumption $\|\beta^\star\|_1 \leq B$ holds automatically with $B = \sqrt{s}/\kappa_0$. Since the latent response has unit variance, the normal equations and the Cauchy–Schwarz inequality give $(\beta^\star)^\top\Sigma_{XX}\beta^\star = (\beta^\star)^\top\Sigma_{XY} \leq \{(\beta^\star)^\top\Sigma_{XX}\beta^\star\}^{1/2}$, so $(\beta^\star)^\top\Sigma_{XX}\beta^\star \leq 1$. The restricted-eigenvalue condition for the s-sparse $\beta^\star$ then gives $\|\beta^\star\|_2 \leq \kappa_0^{-1}$. For fixed κ_0, a bound B independent of s saves only a factor $\sqrt{s}$ in (18).

4.2 Gaussian copula graphical models

Suppose that Σ is invertible and let $\Theta^\star = \Sigma^{-1}$; its off-diagonal zeros are the missing edges of the conditional independence graph of the latent Gaussian vector, hence of X. The jth column $\theta^\star_j$ of $\Theta^\star$

satisfies $\Sigma\theta_j^\star = e_j$, so consider the columnwise program of Liu & Luo (2015) with the normal-score matrix,

$$\widehat{\theta}_j \in \arg\min_{\theta\in\mathbb{R}^p} \left\{ \frac{1}{2}\theta^\top \widehat{\Sigma}^{\mathrm{NS}}\theta - e_j^\top\theta + \lambda \|\theta\|_1 \right\}, \qquad j = 1,\dots,p,$$

and let $\widehat{\Theta}$ have columns $\widehat{\theta}_1,\dots,\widehat{\theta}_p$. If the objective has no minimizer, which can happen when $\widehat{\Sigma}^{\mathrm{NS}}$ is singular, set $\widehat{\theta}_j = 0$; on the event used in Appendix D a minimizer exists. At the target the gradient is $(\widehat{\Sigma}^{\mathrm{NS}} - \Sigma)\theta_j^\star$, so

$$\left\|\widehat{\Sigma}^{\mathrm{NS}}\theta_j^\star - e_j\right\|_\infty \le \left\|\widehat{\Sigma}^{\mathrm{NS}} - \Sigma\right\|_{\max} \left\|\theta_j^\star\right\|_1 . \tag{19}$$

Corollary 4.2 (Gaussian copula graphical model)**.** *Suppose that* $1 \le s \le p$, $\max_j \left\|\theta_j^\star\right\|_0 \le s$, $\max_j \left\|\theta_j^\star\right\|_1 \le M$, $\kappa_\Sigma(s,3) \ge \kappa_0 > 0$ *and* $n \ge A\, s \log p$. *Then the choice* $\lambda = C_\lambda M\sqrt{\log p/n}$, *with* C_λ *as in Corollary 4.1, gives, with probability at least* $1 - C/p$,

$$\max_{j\le p}\left\|\widehat{\theta}_j - \theta_j^\star\right\|_2 \le \frac{CM}{\kappa_0^2}\sqrt{\frac{s\log p}{n}}, \qquad \max_{j\le p}\left\|\widehat{\theta}_j - \theta_j^\star\right\|_1 \le \frac{CMs}{\kappa_0^2}\sqrt{\frac{\log p}{n}}, \tag{20}$$

and consequently

$$\frac{1}{p}\left\|\widehat{\Theta} - \Theta^\star\right\|_F^2 \le CM^2\kappa_0^{-4}\frac{s\log p}{n}.$$

The same three steps apply columnwise: Corollary 2.7 bounds (19) for all j on one event, Corollary 2.4 gives $\kappa_{\widehat{\Sigma}^{\mathrm{NS}}}(s,3) \ge \kappa_0/2$ when $n \ge A\, s\log p$, and the deterministic bound of Appendix D gives (20); the Frobenius bound sums the squared ℓ_2 bounds over columns.

Since $\Theta^\star$ is symmetric, the symmetrization of Liu & Luo (2015), whose (i,j) and (j,i) entries are both set to whichever of $\widehat{\Theta}_{ij}$ and $\widehat{\Theta}_{ji}$ has the smaller absolute value, has max-norm error at most that of $\widehat{\Theta}$. On the event of (20) this error is at most the right side r_n of the first bound there. Setting to zero its entries of absolute value at most r_n then recovers the support of $\Theta^\star$ exactly, provided that every nonzero entry of $\Theta^\star$ exceeds $2r_n$ in absolute value.

5 Discussion

The two obstacles noted in Section 1 are handled at the level of columns rather than entries. The extreme scores are never controlled individually; the object controlled is the Euclidean norm of a whole score-column error, with mean square of order $\log\log n$ (Appendix A). This order reflects the fluctuation of the extreme latent order statistics, not the size of the extreme scores, and the scores need no truncation. The dependence within a column is handled by treating the column as one object: Gaussian concentration applies to the error norm, a Lipschitz function of the whole column. The tools are empirical optimal transport (Bobkov & Ledoux, 2019; Berthet & Fort, 2020), Gaussian concentration and nonasymptotic random matrix theory.

Theorem 2.1 also transfers other oracle conclusions. A quantity whose change under a max-norm perturbation of size $\epsilon \asymp (\log p + \log\log n)/n$ is controlled inherits the oracle bound up to that change. Quadratic forms over unit directions change by at most ϵ times the largest squared ℓ_1 norm among them, which is how Corollary 2.5 compares the two sample matrices; operator norms cost a factor p.

An open question is whether the additive $\log\log n$ can be removed from Theorem 2.1. Appendix A establishes the exact mean-square scale of the chosen coupling, not a lower bound for the oracle discrepancy. The $\log\log n$ term no longer changes the sample-size order in Corollary 2.4.

In conclusion, the second-order oracle approximation of Theorem 2.1 gives the normal-score correlation matrix the two guarantees that sparse estimation in Gaussian copula models requires. Its restricted eigenvalues are controlled at Gaussian-design sample sizes, and its max-norm rate is sharp over the class of correlation matrices. The route, a column-by-column coupling of the score design with the latent Gaussian design, reduces the high-dimensional analysis of normal scores to that of a Gaussian design plus a second-order remainder.

AI use statement

The research question, the main results and the proof strategy are the authors' own. In particular, the central idea of comparing the normal-score matrix directly with the latent Gaussian oracle

(Theorem 2.1), and of deriving both the restricted-eigenvalue and the max-norm guarantees from that single comparison, is the authors'. Under the authors' direction, AI tools (Claude and GPT models) helped identify the probabilistic tools used in the proofs, formulate intermediate lemmas and develop their arguments, draft the proofs and the text, and check references. The authors verified the derivations and restructured the proofs into their present form. The paper contains no data or experiments. The authors take full responsibility for all content.

Reproducibility statement

The paper is theoretical. The sampling model and the global assumptions are stated in Section 2, and every result states its remaining assumptions. Theorem 2.1, Proposition 3.4 and Corollaries 2.3, 2.5 and 2.7 are proved in Section 3; all other results are proved in Appendices A to D.

References

Rina Foygel Barber and Mladen Kolar. ROCKET: Robust confidence intervals via Kendall's tau for transelliptical graphical models. *The Annals of Statistics*, 46(6B):3422–3450, 2018. doi: 10.1214/17-AOS1663.

Philippe Berthet and Jean-Claude Fort. Exact rate of convergence of the expected W_2 distance between the empirical and true Gaussian distribution. *Electronic Journal of Probability*, 25(12): 1–16, 2020. doi: 10.1214/19-EJP410.

Peter J. Bickel, Ya'acov Ritov, and Alexandre B. Tsybakov. Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, 37(4):1705–1732, 2009. doi: 10.1214/08-AOS620.

Sergey G. Bobkov and Michel Ledoux. One-dimensional empirical measures, order statistics, and Kantorovich transport distances. *Memoirs of the American Mathematical Society*, 261(1259), 2019. doi: 10.1090/memo/1259.

T. Tony Cai and Tiefeng Jiang. Phase transition in limiting distributions of coherence of high-dimensional random matrices. *Journal of Multivariate Analysis*, 107:24–39, 2012. doi: 10.1016/j.jmva.2011.11.008. arXiv:1102.2926.

T. Tony Cai and Linjun Zhang. High-dimensional Gaussian copula regression: Adaptive estimation and statistical inference. *Statistica Sinica*, 28(2):963–993, 2018. doi: 10.5705/ss.202016.0041.

Chris A. J. Klaassen and Jon A. Wellner. Efficient estimation in the bivariate normal copula model: Normal margins are least favourable. *Bernoulli*, 3(1):55–77, 1997. doi: 10.2307/3318652.

Christopher Liaw, Abbas Mehrabian, Yaniv Plan, and Roman Vershynin. A simple tool for bounding the deviation of random matrices on geometric sets. In Bo'az Klartag and Emanuel Milman (eds.), *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2014–2016*, volume 2169 of *Lecture Notes in Mathematics*, pp. 277–299. Springer, Cham, 2017. doi: 10.1007/978-3-319-45282-1_18. arXiv:1603.00897.

Han Liu, John Lafferty, and Larry Wasserman. The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research*, 10(80):2295–2328, 2009.

Han Liu, Fang Han, Ming Yuan, John Lafferty, and Larry Wasserman. High-dimensional semiparametric Gaussian copula graphical models. *The Annals of Statistics*, 40(4):2293–2326, 2012. doi: 10.1214/12-AOS1037.

Weidong Liu and Xi Luo. Fast and adaptive sparse precision matrix estimation in high dimensions. *Journal of Multivariate Analysis*, 135:153–162, 2015. doi: 10.1016/j.jmva.2014.11.005.

Ritwik Mitra and Cun-Hui Zhang. Multivariate analysis of nonparametric estimates of large correlation matrices. arXiv:1403.6195, 2014.

Sahand N. Negahban, Pradeep Ravikumar, Martin J. Wainwright, and Bin Yu. A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. *Statistical Science*, 27(4):538–557, 2012. doi: 10.1214/12-STS400.

Garvesh Raskutti, Martin J. Wainwright, and Bin Yu. Restricted eigenvalue properties for correlated Gaussian designs. *Journal of Machine Learning Research*, 11(78):2241–2259, 2010.

Johan Segers, Ramon van den Akker, and Bas J. M. Werker. Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation. *The Annals of Statistics*, 42(5):1911–1940, 2014. doi: 10.1214/14-AOS1244.

Lingzhou Xue and Hui Zou. Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *The Annals of Statistics*, 40(5):2541–2571, 2012. doi: 10.1214/12-AOS1041.

Tuo Zhao, Kathryn Roeder, and Han Liu. Smooth-projected neighborhood pursuit for high-dimensional nonparanormal graph estimation. In *Advances in Neural Information Processing Systems*, volume 25, pp. 162–170. Curran Associates, 2012.

Yue Zhao and Christian Genest. Inference for elliptical copula multivariate response regression models. *Electronic Journal of Statistics*, 13(1):911–984, 2019. doi: 10.1214/19-EJS1534.

A Proofs for Section 3.1

This appendix proves Lemma 3.1. It represents the column error norm as a distance to a fixed set, bounds its mean square through a one-dimensional transport identity, and combines the two by Gaussian concentration. The calculus behind the grid bound comes last.

Unless redefined locally, symbols keep their meaning from the main text throughout the appendices: $\widetilde{Z}$, E and $\mathcal{N}_n$ from Section 3.1; $\widehat{\Sigma}_0^G$, A, B, Δ_0, Δ_B, Ψ_Σ, Ξ and R_n from Section 3.2; P_n, $\widehat{\Sigma}^G$ and $\widehat{R}^G$ from Section 2.1. The letters c and C denote universal constants as in Section 1.

Let $q = (q_1, \dots, q_n)^\top$, and let $\mathcal{Q}_n$ be the orbit of the normalized score vector:

$$\mathcal{Q}_n = \{\Pi q/\sqrt{\gamma_n} : \Pi \text{ is an } n \times n \text{ permutation matrix}\}.$$

Every element of $\mathcal{Q}_n$ has Euclidean norm $\sqrt{n}$.

Lemma A.1 (Projection representation)**.** *For any $z \in \mathbb{R}^n$ with distinct coordinates,*

$$\mathcal{N}_n(z) \in \arg\min_{u \in \mathcal{Q}_n} \|z - u\|_2 . \tag{21}$$

The distance function $g_n(z) := \operatorname{dist}(z, \mathcal{Q}_n)$ is 1-Lipschitz on $\mathbb{R}^n$, and by (21) it equals $\|\mathcal{N}_n(z) - z\|_2$ whenever the coordinates of z are distinct.

Proof of Lemma A.1. Since every vector in $\mathcal{Q}_n$ has the same Euclidean norm, minimizing $\|z - u\|_2^2$ over $\mathcal{Q}_n$ is equivalent to maximizing $z^\top u$. By the rearrangement inequality, the maximum is attained by pairing the scores with the coordinates in the same order, which is exactly $\mathcal{N}_n(z)$. Distance to a fixed nonempty set is 1-Lipschitz. □

The coordinates of a Gaussian column are almost surely distinct, so Lemma A.1 and $\widetilde{Z}_{\cdot j} = \mathcal{N}_n(Z_{\cdot j})$ (Section 3.1) give $\|E_{\cdot j}\|_2 = g_n(Z_{\cdot j})$ almost surely for all j. It remains to bound the mean square of $g_n(Z_{\cdot j})$.

The mean square of one column error is a transport quantity. Let $G = (G_1, \dots, G_n)^\top$ have iid standard normal entries. Define the empirical measure of the Gaussian sample and the uniform measures on the unnormalized and normalized score grids:

$$\mu_n = \frac{1}{n}\sum_i \delta_{G_i}, \qquad \nu_n = \frac{1}{n}\sum_r \delta_{q_r}, \qquad \nu_n^* = \frac{1}{n}\sum_r \delta_{q_r/\sqrt{\gamma_n}}.$$

Both μ_n and ν_n^* have n atoms of mass $1/n$. In one dimension, the monotone coupling is optimal for the quadratic Wasserstein distance W_2. Here it pairs the atoms in increasing order, which is the rank matching that defines $\mathcal{N}_n$. Hence, by Lemma A.1,

$$\frac{1}{n} g_n(G)^2 = W_2^2(\mu_n, \nu_n^*). \tag{22}$$

We bound the right side by passing through $N(0,1)$, using two facts. The deterministic grid is close to $N(0,1)$:

$$W_2^2\{\nu_n, N(0,1)\} \le \frac{12}{n}, \qquad n \ge 2, \tag{23}$$

proved below. The empirical measure of a Gaussian sample satisfies

$$\mathbb{E}W_2^2\{\mu_n, N(0,1)\} \le C\,\frac{\log\log n}{n}, \qquad n \ge 3, \tag{24}$$

(Bobkov & Ledoux, 2019, Corollary 6.14). Berthet & Fort (2020) show that this order is exact. The second fact dominates.

Lemma A.2 (Mean-square coupling error)**.** *There is a universal constant C such that*

$$\mathbb{E}g_n(G)^2 \le C\log\log n, \qquad n \ge 3.$$

Proof of Lemma A.2. We connect μ_n to ν_n^* through $N(0,1)$ and the unnormalized grid ν_n. Monotone pairing gives $W_2(\nu_n, \nu_n^*) = |1 - \sqrt{\gamma_n}|$. Since ν_n and $N(0,1)$ have second moments γ_n and 1, the reverse triangle inequality in L^2 under any coupling gives $|1 - \sqrt{\gamma_n}| \le W_2\{\nu_n, N(0,1)\}$. Hence, by the triangle inequality for W_2,

$$W_2(\mu_n, \nu_n^*) \le W_2\{\mu_n, N(0,1)\} + 2W_2\{\nu_n, N(0,1)\}.$$

Square, use $(x+y)^2 \le 2x^2 + 2y^2$, take expectations and multiply by n. By (22), (23) and (24), $\mathbb{E}g_n(G)^2 \le 2C\log\log n + 96$, and $\log\log n \ge \log\log 3 > 0$ for $n \ge 3$ absorbs the constant term. □

Lemma A.2 is the only place where $\log\log n$ enters the analysis. Lemmas A.1 and A.2 now give Lemma 3.1.

Proof of Lemma 3.1. Each latent column has distribution $N_n(0, I_n)$ and g_n is 1-Lipschitz. Therefore, Gaussian concentration and Jensen's inequality with Lemma A.2 give, for any $x \ge 0$,

$$\mathbb{P}\big\{g_n(Z_{\cdot j}) \ge \sqrt{C\log\log n} + \sqrt{2x}\big\} \le e^{-x}.$$

Take $x = t + \log p$ and apply a union bound over the p columns: the bound fails with probability at most $pe^{-t-\log p} = e^{-t}$, and $\sqrt{2(t+\log p)} \le 2\sqrt{\log p + t}$ gives (11). □

Proof of (23). We bound the interior cells of the quantile grid through a derivative bound for $Q = \Phi^{-1}$ and the extreme cell through a logarithmic integral. Let $\phi = \Phi'$ and $a = \sqrt{\pi/2}$. Since ν_n has n atoms of equal mass $1/n$ and $N(0,1)$ has quantile function Q, the one-dimensional monotone coupling gives

$$W_2^2\left(\nu_n, N(0,1)\right) = \sum_{r=1}^{n} \int_{(r-1)/n}^{r/n} \{Q(u) - q_r\}^2\, du, \qquad q_r = Q\{r/(n+1)\}.$$

Both Q and the grid are antisymmetric about $u = 1/2$, so it suffices to bound the sum over $r \le \lceil n/2 \rceil$ and double it. For odd n the middle cell is then counted twice, which only enlarges the bound.

For $x \ge 0$, the substitution $y = x + z$ gives

$$\frac{\Phi(-x)}{\phi(x)} = \int_0^\infty e^{-xz - z^2/2}\, dz \le \int_0^\infty e^{-z^2/2}\, dz = a.$$

With $x = |Q(u)|$ and $u = \Phi(-x)$ this reads $Q'(u) = 1/\phi(Q(u)) \le a/u$ for $0 < u \le 1/2$, and the symmetry $Q(1-u) = -Q(u)$ then gives

$$Q'(u) \le \frac{a}{u \wedge (1-u)}, \qquad 0 < u < 1. \tag{25}$$

Interior cells. Fix $2 \le r \le \lceil n/2 \rceil$ and $u \in [(r-1)/n, r/n]$. Both u and $r/(n+1)$ lie in this cell, so $|u - r/(n+1)| \le 1/n$, while $u \wedge (1-u) \ge (r-1)/n$ on the cell because $2r \le n+1$. The mean value theorem and (25) give $\{Q(u) - q_r\}^2 \le a^2/(r-1)^2$; integrating over the cell of length $1/n$ and summing over r yields

$$\sum_{r=2}^{\lceil n/2 \rceil} \int_{(r-1)/n}^{r/n} \{Q(u) - q_r\}^2 du \le \frac{a^2}{n} \sum_{k \ge 1} \frac{1}{k^2} = \frac{a^2\pi^2}{6n}.$$

Extreme cell. For $0 < u, v \le 1/2$, integrating (25) gives $|Q(u) - Q(v)| \le a|\log(u/v)|$. Since $n \ge 2$, the extreme cell $(0, 1/n]$ lies in $(0, 1/2]$. Taking $v = 1/(n+1)$ gives

$$\int_0^{1/n} \{Q(u) - q_1\}^2 du \le a^2 \int_0^{1/n} \log^2\{(n+1)u\}\, du$$

$$= \frac{a^2}{n}\left\{2 - 2\log\Big(1+\frac{1}{n}\Big) + \log^2\Big(1+\frac{1}{n}\Big)\right\} \le \frac{2a^2}{n}.$$

The equality uses the substitution $w = (n+1)u$ and $\int_0^b \log^2 w\, dw = b(\log^2 b - 2\log b + 2)$; the final inequality uses $0 < \log(1+1/n) \le 2$.

Adding the two contributions and doubling gives

$$W_2^2\left(\nu_n, N(0,1)\right) \le \frac{2a^2}{n}\left(2 + \frac{\pi^2}{6}\right) = \frac{\pi(2+\pi^2/6)}{n} < \frac{12}{n}, \qquad n \ge 2.$$

□

Remark A.3. Bounding $\{Q(u) - q_1\}^2$ on the extreme cell by $2Q(u)^2 + 2q_1^2$, instead of integrating the logarithm, gives only $O(\log n/n)$. In the proof of Lemma A.2 this term is added to $\mathbb{E}W_2^2(\mu_n, N(0,1)) \asymp \log\log n/n$, so it would dominate and replace $\log\log n$ by $\log n$ in the remainder of Theorem 2.1 and hence of Corollaries 2.3 and 2.7. The two main consequences would not change: $\log n \le 2\sqrt{n}$ keeps (10) intact, and the absorption in Corollary 2.4 works equally well with an additive $\log n$. The refined bound gives the remainders their stated form and makes the grid error negligible against $\log\log n/n$, which the next remark needs.

Remark A.4 (The $\log\log n$ scale is exact). Lemma A.2 is an upper bound, which is all that Theorem 2.1 needs, but the scale it records is the true one. Berthet & Fort (2020) prove

$$\lim_{n\to\infty} \frac{n}{\log\log n}\, \mathbb{E}\, W_2^2(\mu_n, N(0,1)) = 1. \tag{26}$$

This limit concerns $W_2(\mu_n, N(0,1))$ rather than the coupling error $n^{-1}g_n(G)^2 = W_2^2(\mu_n, \nu_n^*)$ of (22). The two are close for every realization: by the triangle inequality and $W_2(\nu_n^*, N(0,1)) \le 2W_2(\nu_n, N(0,1))$, as in the proof of Lemma A.2,

$$|W_2(\mu_n, \nu_n^*) - W_2(\mu_n, N(0,1))| \;\le\; 2W_2(\nu_n, N(0,1))\,.$$

The right side is $O(n^{-1/2})$ by (23), hence $o(\sqrt{\log\log n/n})$. Taking L^2 norms and using (26) gives

$$\mathbb{E}\, g_n(G)^2 \;\sim\; \log\log n \qquad (n\to\infty). \tag{27}$$

The effect comes from the fluctuation of the extreme Gaussian order statistics rather than from their location. The deterministic grid has no fluctuation in its extreme atoms, whereas the empirical measure does.

B Proofs for Section 3.2

This appendix proves Proposition 3.2 and Lemmas 3.3 and 3.5. The first gives the exact decomposition and the conditional Gaussian law; the second bounds the residual cross terms. The third controls the remainder of the oracle expansion using Gaussian moment bounds.

Proof of Proposition 3.2. Let $a_j = A_{jj}$. Each column of $\widetilde{Z}$ has squared norm n, so $(\widehat{\Sigma}^{\mathrm{NS}})_{jj} = 1$, and $\widehat{\Sigma}^{\mathrm{NS}} = \widehat{\Sigma}_0^G + A + A^\top + B$ gives

$$(\widehat{\Sigma}_0^G)_{jj} + 2a_j + B_{jj} = 1, \qquad a_j = -\frac{1}{2}\{(\Delta_0)_{jj} + B_{jj}\}. \tag{28}$$

For fixed j, the Gaussian conditional decomposition of $Z_{\cdot k}$ given $Z_{\cdot j}$ gives

$$Z_{\cdot k} = \Sigma_{kj} Z_{\cdot j} + W_{\cdot k}^{(j)},$$

where $W_{\cdot k}^{(j)}$ is independent of $Z_{\cdot j}$ and has covariance $(1-\Sigma_{jk}^2)I_n$. By the definition of Ξ and this identity,

$$A_{kj} = \Sigma_{kj} a_j + \Xi_{kj}, \qquad A = \Sigma\,\mathrm{Diag}(a_1,\dots,a_p) + \Xi, \qquad \Xi_{kj} = \frac{1}{n}\big(W_{\cdot k}^{(j)}\big)^\top E_{\cdot j}.$$

In particular $\Xi_{jj} = 0$, since $W_{\cdot j}^{(j)} = 0$. Moreover, $E_{\cdot j}$ is a function of $Z_{\cdot j}$, and $W_{\cdot k}^{(j)}$ is independent of $Z_{\cdot j}$. Conditionally on $Z_{\cdot j}$, the last expression is therefore centered Gaussian with variance $n^{-2}(1-\Sigma_{jk}^2)\,\|E_{\cdot j}\|_2^2$, which is (14). Substituting the identity for A and (28) into $\widehat{\Sigma}^{\mathrm{NS}} = \widehat{\Sigma}_0^G + A + A^\top + B$ gives

$$\widehat{\Sigma}^{\mathrm{NS}} - \Sigma = (\widehat{\Sigma}_0^G - \Sigma) - \frac{1}{2}(\Delta_0\Sigma + \Sigma\Delta_0) + R_n.$$

The (j,k) entry of the first two terms is $n^{-1}\sum_i\{Z_{ij}Z_{ik} - \frac{1}{2}\Sigma_{jk}(Z_{ij}^2 + Z_{ik}^2)\} = n^{-1}\sum_i \psi_{jk}(Z_i)$, which gives (13). □

Proof of Lemma 3.3. Set $r = C\{\sqrt{\log\log n} + \sqrt{\log p + t}\}$, the bound in (11), and fix a pair (k, j). By (14), conditionally on $Z_{\cdot j}$ the variable Ξ_{kj} is centered Gaussian with standard deviation at most r/n on the event $\{\|E_{\cdot j}\|_2 \le r\}$, so the Gaussian tail bound at level $t + 2\log p$ gives

$$\mathbb{P}\Big\{|\Xi_{kj}| > \frac{r\sqrt{2(t+2\log p)}}{n},\ \|E_{\cdot j}\|_2 \le r \ \Big|\ Z_{\cdot j}\Big\} \le \frac{2e^{-t}}{p^2}.$$

The event $\{\|E_{\cdot j}\|_2 \le r\}$ is measurable with respect to $Z_{\cdot j}$, so integrating over $Z_{\cdot j}$ gives the same bound unconditionally. A union bound over the p^2 pairs, intersected with the event of (11), gives with probability at least $1 - 3e^{-t}$

$$\|\Xi\|_{\max} \le \frac{r\sqrt{2(t+2\log p)}}{n} \le C\frac{\log p + t + \sqrt{(\log p + t)\log\log n}}{n}.$$

Since $\sqrt{(\log p + t)\log\log n} \le (\log p + t + \log\log n)/2$, this is the stated bound. □

The proof of Lemma 3.5 needs sub-exponential moment bounds for the entries of the latent Gram matrix, which the following representation supplies.

Lemma B.1 (Gaussian representation of the leading term)**.** *Let $Z \sim N_p(0, \Sigma)$ and $\rho = \Sigma_{jk}$. If $|\rho| < 1$, then*

$$U = \frac{Z_j + Z_k}{\sqrt{2(1+\rho)}}, \qquad V = \frac{Z_j - Z_k}{\sqrt{2(1-\rho)}}$$

are independent standard normal variables and

$$\psi_{jk}(Z) = \frac{1-\rho^2}{2}(U^2 - V^2), \qquad Z_j Z_k - \rho = \frac{(1+\rho)(U^2-1) - (1-\rho)(V^2-1)}{2}.$$

If $|\rho| = 1$, then $\psi_{jk}(Z) = 0$ and $Z_j Z_k - \rho = \rho(Z_j^2 - 1)$. In particular the entries of $\Psi_\Sigma(Z)$ and of $ZZ^\top - \Sigma$ are centered and sub-exponential with universal constants.

Proof of Lemma B.1. U and V are standard normal because $\mathrm{Var}(Z_j \pm Z_k) = 2(1 \pm \rho)$, and independent because $\mathrm{Cov}(Z_j + Z_k, Z_j - Z_k) = \mathrm{Var}\, Z_j - \mathrm{Var}\, Z_k = 0$. Solving for Z_j and Z_k gives $Z_j Z_k = \{(1+\rho)U^2 - (1-\rho)V^2\}/2$ and $Z_j^2 + Z_k^2 = (1+\rho)U^2 + (1-\rho)V^2$, and substitution yields both identities. If $|\rho| = 1$ then $Z_k = \rho Z_j$ almost surely, which gives the two stated expressions. Squares of standard normals are sub-exponential, and the coefficients are bounded by 1. □

Proof of Lemma 3.5. We first show that the linearized correlation error differs from the leading term only by a sample-mean correction, and then bound the second-order remainder of the normalization. Let $c > 0$ be a small universal constant, fixed below, and assume $\log p + t \le cn$.

Sample-mean correction. With $\widehat{\Sigma}^G = n^{-1} Z^\top P_n Z$ as in Section 2, let $\Delta = \mathrm{Diag}(\widehat{\Sigma}^G_{11} - 1, \dots, \widehat{\Sigma}^G_{pp} - 1)$ and $L_n = (\widehat{\Sigma}^G - \Sigma) - \frac{1}{2}(\Delta\Sigma + \Sigma\Delta)$, the linearized correlation error. Let $\bar{Z} = n^{-1}\sum_i Z_i$ and $\overline{\Delta} = \mathrm{Diag}(\bar{Z}_1^2, \dots, \bar{Z}_p^2)$. The proof of Proposition 3.2 gives $(\widehat{\Sigma}^G_0 - \Sigma) - \frac{1}{2}(\Delta_0\Sigma + \Sigma\Delta_0) = n^{-1}\sum_i \Psi_\Sigma(Z_i)$. Since $\widehat{\Sigma}^G = \widehat{\Sigma}^G_0 - \bar{Z}\bar{Z}^\top$ and $\Delta = \Delta_0 - \overline{\Delta}$,

$$L_n - \frac{1}{n}\sum_{i=1}^n \Psi_\Sigma(Z_i) = -\bar{Z}\bar{Z}^\top + \frac{1}{2}(\overline{\Delta}\Sigma + \Sigma\overline{\Delta}). \tag{29}$$

Its max norm is at most $2\|\bar{Z}\|_\infty^2$, since $|\Sigma_{jk}| \le 1$. By Lemma B.1, every entry of $\widehat{\Sigma}^G_0 - \Sigma$ is an average of n independent centered sub-exponential variables with universal constants. Bernstein's inequality bounds each such average by $C\{\sqrt{x/n} + x/n\}$ with probability at least $1 - 2e^{-x}$; take $x = t + 2\log p$ and a union bound over the p^2 entries. Each $\bar{Z}_j \sim N(0, 1/n)$, so a Gaussian tail bound and a union bound over j give $\|\bar{Z}\|_\infty \le C\sqrt{(\log p + t)/n}$ with probability at least $1 - e^{-t}$. Since $\|\widehat{\Sigma}^G - \Sigma\|_{\max} \le \|\widehat{\Sigma}^G_0 - \Sigma\|_{\max} + \|\bar{Z}\|_\infty^2$, with probability at least $1 - Ce^{-t}$,

$$u := \|\widehat{\Sigma}^G - \Sigma\|_{\max} \le C\left\{\sqrt{\frac{\log p + t}{n}} + \frac{\log p + t}{n}\right\}, \qquad \|\bar{Z}\|_\infty^2 \le C\frac{\log p + t}{n}.$$

Once c is small enough, $u \le 1/2$.

Normalization remainder. Set $\delta_j = \widehat{\Sigma}^G_{jj} - 1$ and $d_j = (1+\delta_j)^{-1/2}$. Then $|\delta_j| \le u \le 1/2$. The Hessian of $(x,y) \mapsto (1+x)^{-1/2}(1+y)^{-1/2}$ is bounded on $[-1/2, 1/2]^2$, so

$$|d_j d_k - 1 + \tfrac{1}{2}(\delta_j + \delta_k)| \le C(\delta_j^2 + \delta_k^2).$$

For all j, k,

$$\begin{aligned}\widehat{R}^G_{jk} - (\Sigma + L_n)_{jk} &= \widehat{\Sigma}^G_{jk}(d_j d_k - 1) + \frac{\Sigma_{jk}}{2}(\delta_j + \delta_k) \\ &= \widehat{\Sigma}^G_{jk}\{d_j d_k - 1 + \tfrac{1}{2}(\delta_j + \delta_k)\} + \frac{\Sigma_{jk} - \widehat{\Sigma}^G_{jk}}{2}(\delta_j + \delta_k).\end{aligned}$$

Since $\widehat{\Sigma}^G \succeq 0$, $|\widehat{\Sigma}^G_{jk}| \le \sqrt{\widehat{\Sigma}^G_{jj}\widehat{\Sigma}^G_{kk}} \le 1 + u$, and hence

$$\left\| \widehat{R}^G - (\Sigma + L_n) \right\|_{\max} \le Cu^2 \le C\frac{\log p + t}{n}.$$

The last step uses $(\log p + t)/n \le c \le 1$. Combining this bound with (29) and $\|\bar{Z}\|_\infty^2 \le C(\log p + t)/n$ gives the stated bound. □

Remark B.2 (Comparison with the earlier uniform remainder). Zhao & Genest (2019) work in a broader elliptical copula regression framework; their Appendix F concerns the same estimator in the Gaussian copula model (1). Their entrywise remainder bound is, in the notation used here, of order

$$\frac{\log p \, \log^2 n + (\log p)^{3/2}}{n} \qquad \text{under} \qquad \log p = o(n), \quad \log n = O(p)$$

(Zhao & Genest, 2019, Theorem F.1). Relative to the leading scale $\sqrt{\log p / n}$ this leaves the factors $\sqrt{\log p}\,\log^2 n/\sqrt{n}$ and $\log p/\sqrt{n}$. Their hypotheses do not force these to zero: at $\log p = n^{3/4}$ the second diverges like $n^{1/4}$. Both vanish under $\log p = o(\sqrt{n})$. The remainder in (15) is negligible on all of $\log p = o(n)$, with no relation between $\log n$ and p.

C Proofs for Sections 3.3 and 3.4

This appendix proves the two Gaussian-oracle lemmas of Sections 3.3 and 3.4, Corollaries 2.4 and 2.6 and Proposition 2.8. Both lemmas rest on one matrix-deviation inequality for the latent Gaussian design. The corollary proofs use sample-size absorption and a sparse-direction calculation, respectively; the proposition transfers a known limit through Theorem 2.1.

Since only the law of Z enters, we may take $Z = G\Sigma^{1/2}$ with $G \in \mathbb{R}^{n\times p}$ having iid standard normal entries. Liaw et al. (2017, Theorem 1.4 of the arXiv version) prove that for any bounded $\mathcal{T} \subset \mathbb{R}^p$ and any $u \ge 0$, with probability at least $1 - e^{-u^2}$,

$$\sup_{x\in\mathcal{T}} \left| \|Gx\|_2 - \sqrt{n}\,\|x\|_2 \right| \le C\{w(\mathcal{T}) + u\,\mathrm{rad}(\mathcal{T})\}, \tag{30}$$

where $w(\mathcal{T}) = \mathbb{E}\sup_{x\in\mathcal{T}} g^\top x$ for $g \sim N_p(0, I_p)$ and $\mathrm{rad}(\mathcal{T}) = \sup_{x\in\mathcal{T}}\|x\|_2$.

Lemma C.1 (Gaussian design comparison)**.** *For the latent Gaussian design Z and any $t \ge 0$, with probability at least $1 - e^{-t}$, we have*

$$\left| \frac{\|Zv\|_2}{\sqrt{n}} - \sqrt{v^\top \Sigma v} \right| \le C\sqrt{\frac{\log p + t}{n}}\,\|v\|_1 \qquad \textit{simultaneously for all } v \in \mathbb{R}^p.$$

Proof of Lemma C.1. Apply (30) to $\mathcal{T} = \Sigma^{1/2}B_1^p$, the image of the unit ℓ_1 ball. For $x = \Sigma^{1/2}v$, $\|Gx\|_2 = \|Zv\|_2$ and $\|x\|_2 = \sqrt{v^\top\Sigma v}$. Since $\mathrm{diag}(\Sigma) = \mathbf{1}$, $v^\top\Sigma v \le \|v\|_1^2$, so $\mathrm{rad}(\mathcal{T}) \le 1$. By ℓ_1–ℓ_∞ duality, $w(\mathcal{T}) = \mathbb{E}\left\|\Sigma^{1/2}g\right\|_\infty$; every coordinate of $\Sigma^{1/2}g$ is standard normal, so $w(\mathcal{T}) \le \sqrt{2\log(2p)} \le 2\sqrt{\log p}$ for $p \ge 2$. Take $u = \sqrt{t}$. By $2\sqrt{\log p} + \sqrt{t} \le 3\sqrt{\log p + t}$ and homogeneity in v, the stated bound holds with probability at least $1 - e^{-t}$. □

Proof of Lemma 3.6. Fix $t \ge 0$ and set $\delta = \sqrt{(\log p + t)/n}$. Write $Z^c = P_n Z$ and $\bar{Z} = n^{-1}\sum_{i=1}^n Z_i$. Each $\bar{Z}_j \sim N(0, 1/n)$, so a Gaussian tail bound at level $t + 2\log p$ and a union

bound over j give $\|\bar{Z}\|_\infty \le 2\delta$ with probability at least $1-e^{-t}$, since $2p\,e^{-t-2\log p} \le e^{-t}$ for $p \ge 2$. Together with Lemma C.1, with probability at least $1-2e^{-t}$,

$$\left| \frac{\|Zv\|_2}{\sqrt{n}} - \sqrt{v^\top \Sigma v} \right| \le C\delta \|v\|_1 \quad \text{for all } v, \qquad \|\bar{Z}\|_\infty \le 2\delta.$$

As $(I_n - P_n)Zv = \mathbf{1}\,\bar{Z}^\top v$, we have $|\,\|Z^c v\|_2 - \|Zv\|_2\,| \le \sqrt{n}\,|\bar{Z}^\top v| \le 2\sqrt{n}\,\delta\,\|v\|_1$, so on this event

$$\left| \frac{\|Z^c v\|_2}{\sqrt{n}} - \sqrt{v^\top \Sigma v} \right| \le C_0\delta \|v\|_1 \quad \text{for all } v \in \mathbb{R}^p, \qquad C_0 = C + 2.$$

Let $\widehat{\sigma}_j = \|Z^c_{\cdot j}\|_2/\sqrt{n}$. All $\widehat{\sigma}_j$ are positive almost surely, since a zero centered sample variance forces all n observations in that column to coincide. Define $Z^\circ_{\cdot j} = Z^c_{\cdot j}/\widehat{\sigma}_j$, so that $\widehat{R}^G = n^{-1}(Z^\circ)^\top Z^\circ$. Applying the last display to $v = e_j$ gives $\max_j |\widehat{\sigma}_j - 1| \le C_0\delta$. Moreover, the normalization error satisfies the exact identity

$$\frac{\|Z^\circ_{\cdot j} - Z^c_{\cdot j}\|_2}{\sqrt{n}} = |1 - \widehat{\sigma}_j|.$$

Consequently, for all v simultaneously,

$$\begin{aligned}\left|\sqrt{v^\top \widehat{R}^G v} - \sqrt{v^\top \Sigma v}\right| &\le \frac{\|(Z^\circ - Z^c)v\|_2}{\sqrt{n}} + \left| \frac{\|Z^c v\|_2}{\sqrt{n}} - \sqrt{v^\top \Sigma v} \right| \\ &\le \sum_j |v_j| \frac{\|Z^\circ_{\cdot j} - Z^c_{\cdot j}\|_2}{\sqrt{n}} + C_0\delta\|v\|_1 \le 2C_0\delta\|v\|_1 .\end{aligned}$$

The second inequality expands $(Z^\circ - Z^c)v$ over columns and uses the bound for Z^c above; the third uses the identity above and $\max_j |1-\widehat{\sigma}_j| \le C_0\delta$. □

Proof of Lemma 3.7. Work on the event of Lemma 3.6. For $j \ne k$, take $v = e_j + e_k$. Both matrices are correlation matrices, hence $\sqrt{v^\top \widehat{R}^G v} + \sqrt{v^\top \Sigma v} \le 4$. It follows that

$$\begin{aligned}2|\widehat{R}^G_{jk} - \Sigma_{jk}| &= |v^\top(\widehat{R}^G - \Sigma)v| \\ &\le \left|\sqrt{v^\top \widehat{R}^G v} - \sqrt{v^\top \Sigma v}\right| \left(\sqrt{v^\top \widehat{R}^G v} + \sqrt{v^\top \Sigma v}\right) \le 8C\sqrt{\frac{\log p + t}{n}}.\end{aligned}$$

The diagonal differences vanish. This proves the claim after changing the universal constant. □

Proof of Corollary 2.4. The bound (7) is derived in Section 3.3. For the last statement, we first choose K so that the right side of (7) is at most $\eta\kappa_0$ whenever $n \ge Ks(\log p + \log\log n)$, and then show that $n \ge As\log p$ implies this inequality for a suitable $A = A(K)$.

Take $t = \log p$, so that $\log p + t = 2\log p$ and the event of (5) has probability at least $1 - C/p$. Set $K = 2C^2(1+c_0)^2/(\eta^2\kappa_0^2)$, so that $C(1+c_0)\sqrt{2/K} = \eta\kappa_0$. Since $2\log p + \log\log n \le 2(\log p + \log\log n)$, the right side of (7) is at most $\eta\kappa_0 \le \eta\kappa_\Sigma(s, c_0)$ whenever $n \ge Ks(\log p + \log\log n)$. It therefore suffices to show that $n \ge As\log p$ implies $n \ge Ks(\log p + \log\log n)$ for a sufficiently large $A = A(K)$.

Set $m = As\log p$ and take $A \ge 9$, so that $m \ge 9\log 2 > 6$. The map $x \mapsto \log\log x/x$ is decreasing on $[6,\infty)$: its derivative has the sign of $1/\log x - \log\log x$, which is negative there. Hence for $n \ge m$

$$\frac{Ks(\log p + \log\log n)}{n} \le \frac{Ks\log p}{m} + \frac{Ks\log\log m}{m} = \frac{K}{A}\left(1 + \frac{\log\log m}{\log p}\right).$$

The desired implication therefore holds once $A \ge K\{1 + \log\log m/\log p\}$. Since $s \le p$, $\log m \le \log A + 2\log p$, and hence

$$\frac{\log\log m}{\log p} \le \frac{\log(\log A + 2\log p)}{\log p} = g_A(\log p), \qquad g_A(x) := \frac{\log(\log A + 2x)}{x}.$$

Moreover,

$$x^2 g_A'(x) = \frac{2x}{\log A + 2x} - \log(\log A + 2x) < 0 \qquad \text{for } x \ge \log 2.$$

The first term is less than 1 and the second exceeds $\log(\log 9 + 2\log 2) > 1$, so g_A is decreasing and the worst case is $p = 2$. Since $g_A(\log 2)$ grows only like $\log\log A$, the conditions $A \ge 9$ and $A \ge K\{1 + g_A(\log 2)\}$ hold once A is a large enough multiple of $K\{1 + \log\log(e^e + K)\}$. With this A, $n \ge As\log p$ implies $n \ge Ks(\log p + \log\log n)$, and the last statement follows. □

Proof of Corollary 2.6. For v with $\|v\|_2 = 1$ and $\|v\|_0 \le s$, $\|v\|_1 \le \sqrt{s}$ and $v^\top \Sigma v \le \Lambda_s$. Let $a = \sqrt{v^\top \Sigma v}$ and $\hat{a} = \sqrt{v^\top \widehat{\Sigma}^{\mathrm{NS}} v}$; then $|\hat{a} - a| \le r\sqrt{s}$ by (5), and $|\hat{a}^2 - a^2| = |\hat{a} - a|(\hat{a} + a) \le r\sqrt{s}\,(2a + r\sqrt{s})$. Since $a \le \sqrt{\Lambda_s}$, this is the stated bound. □

Proof of Proposition 2.8. Under $\Sigma = I_p$ the columns of Z are independent $N_n(0, I_n)$ vectors, so Cai & Jiang (2012, Corollary 2.1 of the arXiv version) applies to the Pearson correlation matrix $\widehat{R}^G$ and gives $\sqrt{n/\log p}\,\left\|\widehat{R}^G - I_p\right\|_{\max} \to 2$ in probability whenever $\log p = o(n)$. By Theorem 2.1 with $t = \log p$,

$$\sqrt{\frac{n}{\log p}}\,\left\|\widehat{\Sigma}^{\mathrm{NS}} - \widehat{R}^G\right\|_{\max} = O_{\mathbb{P}}\left(\sqrt{\frac{\log p}{n}} + \frac{\log\log n}{\sqrt{n \log p}}\right) = o_{\mathbb{P}}(1).$$

The claim follows from the triangle inequality

$$\left|\left\|\widehat{\Sigma}^{\mathrm{NS}} - I_p\right\|_{\max} - \left\|\widehat{R}^G - I_p\right\|_{\max}\right| \le \left\|\widehat{\Sigma}^{\mathrm{NS}} - \widehat{R}^G\right\|_{\max}.$$

□

D PROOFS FOR THE APPLICATIONS

This appendix proves Corollaries 4.1 and 4.2 through a deterministic error bound for ℓ_1-penalized quadratic programs. Each proof verifies the two inputs of that bound on one event: the gradient condition through Corollary 2.7 and the restricted eigenvalue through Corollary 2.4.

Lemma D.1 (Deterministic error bound)**.** *Let $H \in \mathbb{R}^{p\times p}$ be symmetric positive semidefinite, let $b \in \mathbb{R}^p$, let $\lambda > 0$, and let*

$$\widehat{\theta} \in \arg\min_{\theta\in\mathbb{R}^p} \left\{ \frac{1}{2}\theta^\top H\theta - b^\top\theta + \lambda\,\|\theta\|_1 \right\}.$$

Let $\theta^\star$ have support S with $|S| \le s$, and suppose

$$\|H\theta^\star - b\|_\infty \le \lambda/2. \tag{31}$$

Then $\Delta = \widehat{\theta} - \theta^\star \in \mathcal{C}(S, 3)$, and if $\kappa_H(s, 3) > 0$,

$$\|\Delta\|_2 \le \frac{3\sqrt{s}\,\lambda}{\kappa_H^2(s,3)}, \qquad \|\Delta\|_1 \le \frac{12 s\,\lambda}{\kappa_H^2(s,3)}.$$

Proof. A minimizer exists under (31). Let Q denote the objective. By $H \succeq 0$, Hölder's inequality with (31) and $\|\theta^\star + u\|_1 \ge \|u\|_1 - \|\theta^\star\|_1$, every $u \in \mathbb{R}^p$ satisfies

$$Q(\theta^\star + u) - Q(\theta^\star) \ge -\frac{\lambda}{2}\,\|u\|_1 + \lambda\{\|\theta^\star + u\|_1 - \|\theta^\star\|_1\} \ge \frac{\lambda}{2}\,\|u\|_1 - 2\lambda\,\|\theta^\star\|_1.$$

Hence Q is coercive. Since $Q(\widehat{\theta}) \le Q(\theta^\star)$, expanding the quadratic term gives

$$\frac{1}{2}\Delta^\top H\Delta \le \langle b - H\theta^\star, \Delta\rangle + \lambda(\|\theta^\star\|_1 - \left\|\widehat{\theta}\right\|_1).$$

By (31) and decomposability of the ℓ_1 norm,

$$\frac{1}{2}\Delta^\top H\Delta \le \frac{3}{2}\lambda\,\|\Delta_S\|_1 - \frac{1}{2}\lambda\,\|\Delta_{S^c}\|_1.$$

The left side is nonnegative, so $\|\Delta_{S^c}\|_1 \le 3\,\|\Delta_S\|_1$. On this cone, $\Delta^\top H\Delta \ge \kappa_H^2(s,3)\,\|\Delta\|_2^2$ and $\|\Delta_S\|_1 \le \sqrt{s}\,\|\Delta\|_2$, so $\frac{1}{2}\kappa_H^2(s,3)\,\|\Delta\|_2^2 \le \frac{3}{2}\lambda\sqrt{s}\,\|\Delta\|_2$, which gives the first bound. The second follows from $\|\Delta\|_1 \le 4\,\|\Delta_S\|_1 \le 4\sqrt{s}\,\|\Delta\|_2$. □

Proof of Corollary 4.1. Apply Lemma D.1 with $H = \widehat{\Sigma}^{\mathrm{NS}}_{XX}$, $b = \widehat{\Sigma}^{\mathrm{NS}}_{XY}$ and $\theta^\star = \beta^\star$. Take $t = \log p$ in (10); in dimension $p + 1$, since $\log(p + 1) \le 2\log p$, this gives, with probability at least $1 - C/p$,

$$\left\|\widehat{\Sigma}^{\mathrm{NS}} - \Sigma\right\|_{\max} \le C_{\max}\sqrt{\frac{\log p}{n}}.$$

Together with (17), this verifies (31) for $\lambda = C_\lambda(1 + B)\sqrt{\log p/n}$ when $C_\lambda \ge 2C_{\max}$. Under the sample-size condition of Corollary 4.1, Corollary 2.4 with $c_0 = 3$ and $\eta = 1/2$ gives $\kappa_{\widehat{\Sigma}^{\mathrm{NS}}_{XX}}(s, 3) \ge \frac{1}{2}\kappa_{\Sigma_{XX}}(s, 3) \ge \frac{1}{2}\kappa_0$ with probability at least $1 - C/p$. On the intersection of the two events, which has probability at least $1 - C/p$ after enlarging C, substitution into Lemma D.1 yields (18). □

The restricted-eigenvalue condition also makes $\beta^\star$ the unique s-sparse solution of the normal equations. The difference h of two such solutions is $2s$-sparse and lies in $\mathcal{C}(S,1)$ for S the set of its s largest coordinates in absolute value. It satisfies $h^\top \Sigma_{XX} h = 0$, which $\kappa_{\Sigma_{XX}}(s,3) \geq \kappa_0 > 0$ rules out unless $h = 0$.

Proof of Corollary 4.2. For each j, apply Lemma D.1 with $H = \widehat{\Sigma}^{\mathrm{NS}}$, $b = e_j$, and $\theta^\star = \theta^\star_j$. Taking $t = \log p$ in (10), now in dimension p, again gives $\left\|\widehat{\Sigma}^{\mathrm{NS}} - \Sigma\right\|_{\max} \leq C_{\max}\sqrt{\log p/n}$ with probability at least $1 - C/p$. On that event, (19) verifies (31) simultaneously for all j when $\lambda = C_\lambda M\sqrt{\log p/n}$ and $C_\lambda \geq 2C_{\max}$. Under the sample-size condition of Corollary 4.2, Corollary 2.4 with $c_0 = 3$ and $\eta = 1/2$ gives $\kappa_{\widehat{\Sigma}^{\mathrm{NS}}}(s,3) \geq \frac{1}{2}\kappa_0$ with probability at least $1 - C/p$. On the intersection of the two events, again of probability at least $1 - C/p$, Lemma D.1 gives (20), and summing the squared ℓ_2 bounds over columns gives the Frobenius bound. □