Centered ellipsoid fitting for non-centered Gaussian data

Thresholds, condition numbers, and fitting error

16 October 2026

Abstract

For independent $\mathcal{N}(m_d, I_d)$ points with $\|m_d\| = o(\sqrt{d})$, the threshold for each prescribed condition number and the limiting optimal condition number equal their standard Gaussian values. Without any restriction on the shift, the unrestricted PSD threshold remains 1/4. Below it, the optimal condition number is of order $1 + \|m_d\|/\sqrt{d}$; above it, one event excludes every quadric with nonzero PSD quadratic part, whatever its linear and constant terms. At every positive density the optimal normalized squared fitting error has the unshifted Gaussian limit. These centered-fitting statistics also apply exactly to every mixture of $\mathcal{N}(m_d, I_d)$ and $\mathcal{N}(-m_d, I_d)$.

The nonexistence proof corrects a dual certificate using independent reserve points to annihilate constants and linear functions. It extends to centrally symmetric admissible mixtures of product laws. For these mixtures, positive definite fits below the unshifted threshold also exist for arbitrary shifts supported on $o(d)$ coordinates. The general mixture inputs and the conclusions of this note are proved in Rocq. The conclusions are seven public propositions, each proved by a theorem; their statements are frozen with the repository's checker, and the formalization is released under the tag `shifted`. Appendix A names them and records several differences between the proofs, including a sparse-shift existence theorem that does not require symmetry. In particular, two supporting assertions of the text are not formalized, and the formal proof needs neither: the almost-sure invertibility of the reserve Gram matrix in Lemma 6.2 and the constant of Tropp's matrix Rademacher bound in (42). No claim of novelty is made.

1 Statement and scope

For a sample $X = (X_1, \dots, X_n)$ in $\mathbb{R}^d$, let

$$\mathcal{F}_{n,d}(C) = \{X : \exists Q \succ 0, \quad X_i^{\mathsf{T}} Q X_i = 1 \ (1 \le i \le n), \quad \operatorname{cond}(Q) \le C\},$$

where $\operatorname{cond}(Q) = \lambda_{\max}(Q)/\lambda_{\min}(Q)$, and set

$$\mathcal{C}(X) = \inf\{C \ge 1 : X \in \mathcal{F}_{n,d}(C)\}, \qquad \inf \varnothing = +\infty.$$

The fitting events and $\mathcal{C}$ are Borel measurable; the precise formal sources are listed in Appendix A. The ellipsoids remain centered at the origin.

Write $\alpha_C = \alpha(C, 3)$ for the Gaussian threshold. For completeness, with

$$\nu(dz) = \pi^{-1}\sqrt{2 - z^2}\,\mathbf{1}_{[-\sqrt{2},\sqrt{2}]}(z)\,dz, \qquad m(t) = \int (z - t)_+ \,\nu(dz), \quad v(t) = \int (z - t)_+^2 \,\nu(dz),$$

it is given by

$$m(-t_C) = C\, m(C t_C), \qquad 0 < t_C < \sqrt{2}/C, \qquad \alpha_C = \tfrac{1}{2} - v(-t_C) - v(C t_C). \tag{1}$$

The root is unique, and $C \mapsto \alpha_C$ is continuous and strictly increasing from $(1, \infty)$ onto $(0, 1/4)$. Its inverse is denoted by $C_\star(\alpha) = C_\star(\alpha, 3)$. Appendix A gives the formal sources for the root, the curve properties and the inverse.

Theorem 1.1 (Non-centered Gaussian data)**.** *Let $m_d \in \mathbb{R}^d$ be deterministic and put $a_d = \|m_d\|$. Suppose*

$$\frac{a_d}{\sqrt{d}} \longrightarrow 0, \qquad \frac{n(d)}{d^2} \longrightarrow \alpha > 0. \tag{2}$$

Let $X_i = Z_i + m_d$, where the Z_i are independent $\mathcal{N}(0, I_d)$ vectors. For every fixed $C > 1$,

$$\mathbb{P}\{X \in \mathcal{F}_{n(d),d}(C)\} \longrightarrow \begin{cases} 1, & \alpha < \alpha_C, \\ 0, & \alpha > \alpha_C. \end{cases} \tag{3}$$

If $0 < \alpha < 1/4$, then

$$\mathcal{C}(X) \xrightarrow{\mathbb{P}} C_\star(\alpha). \tag{4}$$

The point law has mean m_d and covariance I_d, so it need not satisfy the centering assumption of the existing mixture theorem. Since $\alpha_C \uparrow 1/4$, the result gives positive definite fits at every $0 < \alpha < 1/4$ by taking C large enough. Theorem 1.3 also gives bounded condition numbers for shifts of order $\sqrt{d}$ and matching growth bounds for larger shifts. It does not identify the precise condition-number curve when $a_d/\sqrt{d}$ has a nonzero limit. No assertion is made at $\alpha = \alpha_C$.

Remark 1.2 (The sign mixture gives the same fitting problem). For any $w_d \in [0, 1]$, the conclusions of Theorem 1.1 also hold for independent points with law

$$w_d\mathcal{N}(m_d, I_d) + (1 - w_d)\mathcal{N}(-m_d, I_d).$$

In fact, the equality of fitting statistics is exact at every n, d. Write these points as $\widetilde{X}_i = Z_i + \sigma_i m_d$, with independent signs satisfying $\mathbb{P}(\sigma_i = 1) = w_d$, independent of the Z_i. Then $Y_i = \sigma_i\widetilde{X}_i = \sigma_i Z_i + m_d$ are independent $\mathcal{N}(m_d, I_d)$ points, while $Y_iY_i^{\mathsf{T}} = \widetilde{X}_i\widetilde{X}_i^{\mathsf{T}}$. Thus the entire sets of fitting matrices coincide samplewise, as do their optimal condition numbers and normalized fitting errors. This includes $w_d = 0, 1$ and does not assert equality of the point distributions. The identity applies to positive semidefinite as well as positive definite fitting matrices and requires no bound on $\|m_d\|$.

Theorem 1.3 (The unrestricted threshold for arbitrary shifts)**.** *Let $n(d)/d^2 \to \alpha \in (0, \infty)$, let $m_d \in \mathbb{R}^d$ be arbitrary and deterministic, and let $X_i = Z_i + m_d$ for independent $Z_i \sim \mathcal{N}(0, I_d)$. There is no restriction on $\|m_d\|$.*

1. *If $\alpha < 1/4$, then with probability tending to one there is $Q \succ 0$ such that $X_i^{\mathsf{T}}QX_i = 1$ for every i. More precisely, there are constants $0 < c_\alpha \le C_\alpha < \infty$ such that, with probability tending to one,*

$$c_\alpha\left(1 + \frac{\|m_d\|}{\sqrt{d}}\right) \le \mathcal{C}(X) \le C_\alpha\left(1 + \frac{\|m_d\|}{\sqrt{d}}\right). \tag{5}$$

2. *If $\alpha > 1/4$, then with probability tending to one there is no $Q \succeq 0$ such that $X_i^{\mathsf{T}}QX_i = 1$ for every i. More strongly, there are Borel sets Ω_d of Gaussian samples with $\mathbb{P}(\Omega_d) \to 1$ such that, on Ω_d, no $Q \succeq 0$, $Q \neq 0$, $\ell \in \mathbb{R}^d$, and $c \in \mathbb{R}$ satisfy*

$$Z_i^{\mathsf{T}}QZ_i + 2\ell^{\mathsf{T}}Z_i + c = 0 \qquad (1 \le i \le n). \tag{6}$$

   In particular, no PSD fit exists for $Z_i + m$, simultaneously for every $m \in \mathbb{R}^d$. No ellipsoid, whatever its center, passes through all the points.

The existence and nonexistence conclusions for centered fits, and the condition-number bounds, also hold for independent points with law $w_d\mathcal{N}(m_d, I_d) + (1 - w_d)\mathcal{N}(-m_d, I_d)$, for every $w_d \in [0, 1]$.

Section 6 proves this theorem and identifies the limiting fitting error in Theorem 6.3. Section 7 extends the unrestricted nonexistence argument to symmetric non-Gaussian mixtures. Exact fitting at the critical density, general component means and covariances, and existence for dense shifts of non-Gaussian data are not settled here.

2 Inputs and deterministic perturbation

Here $\mathbb{S}_p$ denotes the real symmetric $p \times p$ matrices, $\|\cdot\|$ is the Euclidean norm, and logarithms are natural. The general mixture condition-number and interpolation theorems are proved in Rocq. We use their elementary Gaussian specializations. For a standard normal scalar G_0,

$$\mathbb{E}G_0 = 0, \qquad \mathbb{E}G_0^2 = 1, \qquad \mathbb{E}G_0^4 = 3, \qquad \mathbb{E}e^{G_0^2/4} = \sqrt{2} \le 2. \tag{7}$$

Hence the one-component Gaussian family is admissible with $K = 2$ and fourth moment three. The Rocq development proves these scalar facts and exports the Gaussian specializations used below, in `GaussianInputs.v` and `GaussianExports.v` of `extension/Shifted/`; see Appendix A.

Gaussian fitting. If $g_1, \dots, g_N$ are independent $\mathcal{N}(0, I_p)$ and $N/p^2 \to \beta > 0$, then for fixed $D > 1$ the probability of $\mathcal{F}_{N,p}(D)$ tends to one for $\beta < \alpha_D$, and to zero for $\beta > \alpha_D$. This follows by applying `condition_statement_holds` to the family just described.

Simultaneous interpolation. If $0 < \beta < 1/2$ and $N/p^2 \to \beta$, there is $L_\beta > 0$ such that, outside Borel exceptional sets of probability tending to zero, the Gaussian sample has the following property: for every $y \in \mathbb{R}^N$ there exist $\Delta \in \mathbb{S}_p$ and $b \in \mathbb{R}$ with

$$g_i^{\mathsf{T}} \Delta g_i - \operatorname{tr} \Delta = \sqrt{p}\,(y_i + b) \qquad (1 \le i \le N), \tag{8}$$

$$\|\Delta\|_{\mathrm{op}} \le L_\beta \left(\frac{\|y\|_2}{p} + \frac{(\log p)^4}{\sqrt{p}} \|y\|_\infty \right). \tag{9}$$

This is the Gaussian specialization of `mixture_interpolation_holds` in `extension/Mixture/Interpolation.v`, obtained from `mixture_small_residual`; its source is the interpolation theorem of Koehler–Sohn [1, Theorem 1.9]. The formal statement gives an exceptional set of probability at most each prescribed $\epsilon > 0$ in every sufficiently large dimension. A diagonal choice gives the sets used here.

The conclusion holds simultaneously for every y. We use it on the same sample as the Gaussian fitting event. All probability bounds use the Borel exceptional sets, without asserting measurability of the full interpolation-property event or choosing a measurable correction matrix.

Lemma 2.1 (Stability under a small correction)**.** *Suppose $A \succ 0$, $\operatorname{tr} A = p$, and $\operatorname{cond}(A) \le D$. Then $\lambda_{\min}(A) \ge 1/D$ and $\|A\|_{\mathrm{op}} \le D$. If E is symmetric and $\|E\|_{\mathrm{op}} \le \delta < 1/D$, then $A + E \succ 0$ and*

$$\operatorname{cond}(A + E) \le D + \frac{(D+1)\delta}{1/D - \delta}. \tag{10}$$

Also, $\operatorname{diag}(1, A)$ *has trace $p + 1$ and condition number at most D.*

Proof. The mean eigenvalue of A is one, so $\lambda_{\min}(A) \le 1 \le \lambda_{\max}(A)$ and the assertions about A and $\operatorname{diag}(1, A)$ follow. Writing $\ell = \lambda_{\min}(A)$, the extremal eigenvalue bounds give

$$\operatorname{cond}(A + E) \le \frac{D\ell + \delta}{\ell - \delta} = D + \frac{(D+1)\delta}{\ell - \delta} \le D + \frac{(D+1)\delta}{1/D - \delta}. \qquad \square$$

We will repeatedly rescale a fitting matrix to have trace equal to the dimension. Its quadratic values then equal some common level $q > 0$, instead of one. Conversely, a positive definite matrix with a common quadratic value on nonzero observations has a positive common level; division by that level gives an exact fit without changing its condition number. This observation removes any need to estimate the scalar b in (8).

3 Existence below the threshold

Assume $\alpha < \alpha_C$ and choose $1 < C_- < C$ with $\alpha < \alpha_{C_-}$. Rotate m_d to $a_d e_1$ and write

$$X_i = (u_i, g_i), \qquad u_i = \xi_i + a_d, \qquad p = d - 1,$$

where the ξ_i are standard Gaussian, the g_i are $\mathcal{N}(0, I_p)$, and all these variables are independent. This uses Gaussian rotation invariance and splitting off the first coordinate. Since $n(d)/p^2 \to \alpha$, Gaussian fitting gives, with probability tending to one, a matrix $A \succ 0$ such that

$$\operatorname{tr} A = p, \qquad \operatorname{cond}(A) \le C_-, \qquad g_i^{\mathsf{T}} A g_i = q \quad \text{for every } i.$$

Interpolate on this same Gaussian sample the centered targets

$$y_i = -\frac{u_i^2 - (1 + a_d^2)}{\sqrt{p}}. \tag{11}$$

Here

$$u_i^2 - (1 + a_d^2) = \xi_i^2 - 1 + 2a_d\xi_i, \qquad \operatorname{Var}(u_i^2) = 2 + 4a_d^2.$$

Since $n(d) = O(d^2)$, the second-moment bound and the Gaussian maximum bound give

$$\frac{\|y\|_2}{p} = O_{\mathbb{P}}\left(\frac{1 + a_d}{\sqrt{d}}\right), \tag{12}$$

$$\|y\|_\infty = O_{\mathbb{P}}\left(\frac{a_d\sqrt{\log d} + \log d}{\sqrt{d}}\right). \tag{13}$$

For example, $\mathbb{E}\|y\|_2^2 = n(d)(2 + 4a_d^2)/p$ proves the first estimate by Markov's inequality, and $\max_i |\xi_i| = O_{\mathbb{P}}(\sqrt{\log d})$ proves the second. Because $\alpha < 1/4 < 1/2$, interpolation applies. Outside its Borel exceptional set, there exists a correction with

$$\|\Delta\|_{\mathrm{op}} \le S_d := L_\alpha\left(\frac{\|y\|_2}{p} + \frac{(\log p)^4}{\sqrt{p}}\|y\|_\infty\right).$$

This upper bound is a measurable random variable, and

$$S_d = O_{\mathbb{P}}\left(\frac{1 + a_d}{\sqrt{d}} + \frac{a_d(\log d)^{9/2}}{d} + \frac{(\log d)^5}{d}\right) = o_{\mathbb{P}}(1) \tag{14}$$

whenever $a_d = o(\sqrt{d})$.

Set $Q = \operatorname{diag}(1, A + \Delta)$ and choose

$$\delta = \frac{C - C_-}{2C_-(C + 1)} < \frac{1}{C_-}.$$

Then $C_- + (C_- + 1)\delta/(1/C_- - \delta) \le C$. On $\{S_d \le \delta\}$, Lemma 2.1, applied to $\operatorname{diag}(1, A)$, gives $Q \succ 0$ and $\operatorname{cond}(Q) \le C$. Moreover,

$$\begin{aligned} X_i^{\mathsf{T}} Q X_i &= u_i^2 + q + \operatorname{tr}\Delta + \sqrt{p}\,(y_i + b) \\ &= q + 1 + a_d^2 + \operatorname{tr}\Delta + \sqrt{p}\,b. \end{aligned}$$

This is independent of i. It is positive because $Q \succ 0$ and the observations are almost surely nonzero. Rescaling Q gives an exact fit of condition number at most C. The Gaussian fitting event, the complement of the interpolation exceptional set, and $\{S_d \le \delta\}$ have intersection of probability tending to one. This proves the first half of (3).

4 A uniform residual gap for Gaussian points

The missing quantitative statement is contained in the proof of `condition_unsat_gap`, rather than in its exported conclusion. We extract it here, retaining the squared loss that the exact-fit proof sets to zero. This step concerns the actual quadratic features of Gaussian points, not just auxiliary Gaussian matrix measurements.

Write

$$\mathcal{K}_d(C) = \{R \in \mathbb{S}_d : R \succ 0,\ \operatorname{tr} R = d,\ \operatorname{cond}(R) \le C\}, \qquad W_i = \frac{Z_i Z_i^{\mathsf{T}} - I_d}{\sqrt{d}}.$$

For a tuple $U = (U_1, \dots, U_n)$ of symmetric matrices define

$$\mathcal{E}_U(R, b) = \frac{1}{n}\sum_{i=1}^{n}(b - \langle U_i, R\rangle)^2, \qquad \langle A, B\rangle = \operatorname{tr}(AB).$$

Let P_n be orthogonal projection onto the vectors in $\mathbb{R}^n$ with coordinate sum zero. For any point sample Z, set

$$\mathcal{L}_Z(R) = \frac{1}{\sqrt{nd}}\, P_n((Z_i^{\mathsf{T}} R Z_i)_{i=1}^n). \tag{15}$$

Subtracting $\operatorname{tr} R$ from every quadratic value has no effect under P_n, so

$$\|\mathcal{L}_Z(R)\|_2^2 = \min_{b\in\mathbb{R}} \mathcal{E}_W(R, b). \tag{16}$$

Lemma 4.1 (Extracted data gap)**.** *Let $C > 1$, $\alpha > \alpha_C$, and $n(d)/d^2 \to \alpha$. Choose $C_p > C$ with $\alpha_{C_p} < \alpha$, and put*

$$g = \frac{(\sqrt{\alpha} - \sqrt{\alpha_{C_p}})^2}{64\alpha} > 0. \tag{17}$$

For independent $Z_i \sim \mathcal{N}(0, I_d)$ there are Borel sets Ω_d with $\mathbb{P}(\Omega_d) \to 1$ such that, on Ω_d,

$$\mathcal{E}_W(R, b) \ge g/4 \qquad \text{for every } R \in \mathcal{K}_d(C),\ b \in \mathbb{R}. \tag{18}$$

In particular, $\inf_{R\in\mathcal{K}_d(C)} \|\mathcal{L}_Z(R)\|_2^2 \ge g/4$ *there.*

Proof. We spell out the estimates from `extension/Condition/Unsat.v` that make the assertion uniform in R and retain a positive residual. The Gaussian specialization is justified by (7).

Cover and regularized objectives. The finite cover `cond_cover` supplies numbers $0 < \ell_j < 1 < h_j < C_p \ell_j$, $1 \le j \le J$, and a barrier bound B, independent of d. Enlarge B to $\max(B, 0)$. Every $R \in \mathcal{K}_d(C)$ lies in some slice

$$\mathcal{D}_{j,d} = \{S \in \mathbb{S}_d : \operatorname{tr} S = d,\ \ell_j I_d \prec S \prec h_j I_d\}$$

with $\mathcal{B}_{j,d}(R) \le dB$. Here the normalized barrier is

$$\mathcal{B}_{j,d}(S) = -\log\det(S - \ell_j I_d) - \log\det(h_j I_d - S) + d\log((1 - \ell_j)(h_j - 1)).$$

It is nonnegative on the slice. Let λ_d be the operator norm of the covariance of W_i, so that $0 < d\lambda_d \le 2$, and put $\gamma_d = (d\sqrt{d})^{-1}$. For positive weights β_j chosen below, the slice value used in the formal proof is

$$\Phi_j(U) = \inf_{\substack{S\in\mathcal{D}_{j,d}\\ b\in\mathbb{R}}} \left\{\frac{1}{2}\mathcal{E}_U(S, b) + \beta_j(\lambda_d\|S\|_F^2 + b^2) + \gamma_d \mathcal{B}_{j,d}(S)\right\}. \tag{19}$$

The norm $\|\cdot\|_F$ is Frobenius norm. This is `af_value` at fourth moment three.

The simultaneous lower bounds. For auxiliary independent centered Gaussian matrix measurements H_i with the same covariance as W_i, `gaussian_gap` gives $\mathcal{E}_H(S, b) \geq g$ throughout $\mathcal{K}_d(C_p)$ outside an event of probability tending to zero. Each $\mathcal{D}_{j,d}$ is contained in that cone slice. Nonnegativity of the ridge and barrier therefore gives $\Phi_j(H) \geq g/2$ on this event, and $\mathbb{E}\Phi_j(H) \geq g/4$ for all large d. The comparison and concentration estimates used in `condition_unsat_gap` then yield, simultaneously for all j,

$$\Phi_j(W) \geq g/4 - \beta_j/4 \tag{20}$$

outside Borel exceptional sets of probability tending to zero. These are precisely `af_center_ge`, `af_conc`, `sl_err_small`, and `sl_prob_small`, applied to the finitely many fixed slices. None of these events assumes an exact fit.

The minimizing intercept and the upper bound. On the Borel a priori event of `mixture_adjoint_bound`, for a constant C_a and every $z \in \mathbb{R}^n$,

$$\left\| \sum_i z_i W_i \right\|_F \leq C_a \sqrt{d}\, \|z\|_2.$$

For each $R \in \mathcal{K}_d(C)$ choose a covering slice j and set

$$b_R = \frac{1}{n} \sum_i \langle W_i, R \rangle = \langle \overline{W}, R \rangle, \qquad \overline{W} = \frac{1}{n} \sum_i W_i.$$

The a priori bound and $\|R\|_F^2 \leq h_j d$ give, eventually,

$$|b_R| \leq C_a \sqrt{d/n}\, \|R\|_F, \qquad b_R^2 \leq c_b h_j, \qquad c_b = 2C_a^2/\alpha, \tag{21}$$

using $n \geq \alpha d^2/2$. Unlike `common_value_le`, this argument does not assume equal feature values: b_R is their empirical mean by definition.

Use the same parameter choices as the formal proof:

$$\theta = \frac{g}{16(1 + c_b)}, \qquad \beta_j = \frac{\theta}{2(2h_j + 1)}.$$

Then

$$\beta_j \leq g/32, \qquad \beta_j(2h_j + b_R^2) \leq g/32, \qquad \gamma_d \mathcal{B}_{j,d}(R) \leq B/\sqrt{d} \leq g/32 \tag{22}$$

for all sufficiently large d. Evaluating (19) at (R, b_R), and using $d\lambda_d \leq 2$, gives

$$\Phi_j(W) \leq \tfrac{1}{2} \mathcal{E}_W(R, b_R) + g/16. \tag{23}$$

This is exactly the proof of `af_value_common` with its loss term retained instead of set to zero.

Combining (20)–(23),

$$\tfrac{1}{2} \mathcal{E}_W(R, b_R) + g/16 \geq g/4 - g/128 = 31g/128, \qquad \mathcal{E}_W(R, b_R) \geq 23g/64 > g/4.$$

Finally, $\mathcal{E}_W(R, b) = \mathcal{E}_W(R, b_R) + (b - b_R)^2$, so the bound holds for every b. The event Ω_d is the intersection of the a priori event and the complements of the finitely many slice exceptional sets, with the same diagonal choice as in Section 2. It is Borel and has probability tending to one, independently of R and b. No restriction $\alpha < 1/2$ was used. □

5 Nonexistence and the optimal condition number

Assume $\alpha > \alpha_C$ and use the full sample of $n = n(d)$ points. Let G have rows Z_i^T. The standard Gaussian matrix norm bound gives $\|G\|_{\mathrm{op}} = O_{\mathbb{P}}(\sqrt{n} + \sqrt{d}) = O_{\mathbb{P}}(d)$. Expanding $X_i = Z_i + m_d$ and applying P_n removes the common term $m_d^\mathsf{T} R m_d$. Uniformly over $R \in \mathcal{K}_d(C)$,

$$\|\mathcal{L}_X(R) - \mathcal{L}_Z(R)\|_2 = \frac{2\|P_n G R m_d\|_2}{\sqrt{nd}} \le T_d := \frac{2C a_d \|G\|_{\mathrm{op}}}{\sqrt{nd}} = O_{\mathbb{P}}\left(\frac{a_d}{\sqrt{d}}\right) = o_{\mathbb{P}}(1). \tag{24}$$

Here $\|R\|_{\mathrm{op}} \le C$ by Lemma 2.1, and T_d is a measurable bound independent of the choice of R.

A hypothetical fit of X with condition number at most C, rescaled to trace d, belongs to $\mathcal{K}_d(C)$ and satisfies $\mathcal{L}_X(R) = 0$. On Ω_d from Lemma 4.1, it would therefore imply $\sqrt{g}/2 \le T_d$. Consequently,

$$\mathbb{P}\{X \in \mathcal{F}_{n,d}(C)\} \le \mathbb{P}(\Omega_d^c) + \mathbb{P}\{T_d \ge \sqrt{g}/2\} \longrightarrow 0.$$

This proves nonexistence without subsampling or reverse interpolation. Uniformity in (24) handles a matrix selected from the data; no measurable choice of a fitting matrix is needed.

For $0 < \alpha < 1/4$, the optimum now follows from the threshold and monotonicity of $\mathcal{F}_{n,d}(C)$ in C. Given $\eta > 0$, existence at $C_\star(\alpha) + \eta/2$ gives $\mathbb{P}\{\mathcal{C}(X) > C_\star(\alpha) + \eta\} \to 0$. If $C_\star(\alpha) - \eta \le 1$, the lower deviation is impossible. Otherwise choose $D \in (C_\star(\alpha) - \eta, C_\star(\alpha))$. The definition of an infimum and monotonicity give

$$\{\mathcal{C}(X) < C_\star(\alpha) - \eta\} \subseteq \{X \in \mathcal{F}_{n,d}(D)\}.$$

The right-hand probability tends to zero because $\alpha > \alpha_D$. This proves (4) without assuming attainment and completes Theorem 1.1.

6 Unrestricted fits for arbitrary Gaussian shifts

The bound in (24) uses a fixed condition number. We now prove Theorem 1.3 by a different argument. Above $1/4$ we construct a positive definite dual certificate whose weights annihilate both constants and linear functions. Such a certificate is unchanged by translation of the observations.

6.1 Positive definite fits below $1/4$

Suppose $0 < \alpha < 1/4$ and choose a fixed $D > 1$ with $\alpha < \alpha_D$. As in Section 3, rotate the shift to $a_d e_1$ and write $X_i = (\xi_i + a_d, g_i)$, with $p = d - 1$. With probability tending to one, the Gaussian sample (g_i) admits $A \succ 0$ with $\operatorname{tr} A = p$, $\operatorname{cond}(A) \le D$, and $g_i^\mathsf{T} A g_i = q$ independent of i.

Fix a sufficiently small $\varepsilon > 0$, depending only on α and D, and give the first coordinate the positive weight

$$s_d = \min\left(1, \frac{\varepsilon\sqrt{d}}{a_d}\right), \qquad s_d = 1 \quad \text{if } a_d = 0.$$

Interpolate the targets

$$y_i = -\frac{s_d((\xi_i + a_d)^2 - (1 + a_d^2))}{\sqrt{p}}.$$

Since $s_d \le 1$ and $s_d a_d \le \varepsilon\sqrt{d}$, the law of large numbers for ξ_i^2 and $(\xi_i^2 - 1)^2$ gives

$$\frac{\|y\|_2}{p} \le \frac{s_d}{p^{3/2}}\|(\xi_i^2 - 1)_i\|_2 + \frac{2 s_d a_d}{p^{3/2}}\|(\xi_i)_i\|_2 \le 2\sqrt{\alpha}\,\varepsilon + o_{\mathbb{P}}(1), \tag{25}$$

$$\frac{(\log p)^4}{\sqrt{p}}\|y\|_\infty = O_{\mathbb{P}}\left(\frac{(\log d)^5}{d} + \frac{\varepsilon(\log d)^{9/2}}{\sqrt{d}}\right) = o_{\mathbb{P}}(1). \tag{26}$$

The second estimate uses $\max_i |\xi_i| = O_{\mathbb{P}}(\sqrt{\log d})$. In particular, the first bound has a deterministic leading constant; a tightness estimate alone would not suffice for this choice of s_d. Simultaneous interpolation therefore supplies a symmetric correction Δ and an intercept b_0 with the measurable upper bound

$$\|\Delta\|_{\mathrm{op}} \le S'_d := L_\alpha \left(\frac{\|y\|_2}{p} + \frac{(\log p)^4}{\sqrt{p}} \|y\|_\infty \right) \le 2L_\alpha \sqrt{\alpha}\, \varepsilon + o_{\mathbb{P}}(1).$$

Choose $0 < \varepsilon \le 1$ so that $2L_\alpha \sqrt{\alpha}\, \varepsilon < 1/(4D)$. With probability tending to one, $\|\Delta\|_{\mathrm{op}} \le 1/(2D)$, and hence

$$(2D)^{-1} I_p \preceq A + \Delta \preceq 2D I_p.$$

The matrix $Q = \mathrm{diag}(s_d, A + \Delta)$ is then positive definite and

$$X_i^{\mathsf{T}} Q X_i = q + s_d(1 + a_d^2) + \operatorname{tr} \Delta + \sqrt{p}\, b_0$$

is independent of i. Its common value is positive, so rescaling gives an exact positive definite fit, with

$$\mathrm{cond}(Q) \le \frac{4D^2}{s_d} \le \frac{4D^2}{\varepsilon} \left(1 + \frac{a_d}{\sqrt{d}} \right). \tag{27}$$

The Borel exceptional-set argument is the same as in Section 3. This proves existence and the upper bound in (5), without a growth restriction on a_d.

A matching lower bound. For any fixed density $\alpha > 0$, let G have rows Z_i^{T}. The adjoint bound used in (38), by duality, gives on events of probability tending to one a fixed $L > 0$, depending only on α, such that

$$\|(Z_i^{\mathsf{T}} Q Z_i - \operatorname{tr} Q)_i\|_2 \le Ld\, \|Q\|_F \qquad (Q \in \mathbb{S}_d).$$

Also $\sigma_{\min}(P_n G)/\sqrt{n} \to 1$ in probability: an orthogonal change of the row coordinates identifies its nonzero rows with an $(n-1) \times d$ standard Gaussian matrix. If $Q \succ 0$ fits $Z + m$, centering its fit equations gives

$$P_n (Z_i^{\mathsf{T}} Q Z_i)_i = -2 P_n G Q m.$$

Consequently, on the same events, simultaneously for every such m, Q,

$$\begin{aligned} 2\sigma_{\min}(P_n G) \lambda_{\min}(Q) \|m\| &\le 2 \|P_n G Q m\|_2 \\ &\le Ld \|Q\|_F \le Ld^{3/2} \lambda_{\max}(Q). \end{aligned}$$

Canceling $\lambda_{\min}(Q) > 0$ gives the scale-invariant bound

$$\mathrm{cond}(Q) \ge \frac{2\sigma_{\min}(P_n G)}{Ld} \frac{\|m\|}{\sqrt{d}} = \left(\frac{2\sqrt{\alpha}}{L} + o_{\mathbb{P}}(1) \right) \frac{\|m\|}{\sqrt{d}}. \tag{28}$$

The random coefficient is independent of m and Q. Together with $\mathrm{cond}(Q) \ge 1$, this proves the lower bound in (5) after decreasing a deterministic positive constant c_α. Taking the infimum over fitting matrices requires no attainment assumption. In particular, for $\alpha < 1/4$, shifts of order $\sqrt{d}$ admit bounded condition numbers, whereas $a_d/\sqrt{d} \to \infty$ forces the optimal condition number to diverge linearly in $a_d/\sqrt{d}$, up to fixed multiplicative constants.

6.2 A Gaussian certificate with a positive margin

For a sample $Z = (Z_1, \dots, Z_N)$ define its unrestricted normalized squared residual by

$$\Gamma_{N,d}(Z) = \min_{\substack{R \succeq 0 \\ \operatorname{tr} R = d}} \|\mathcal{L}_Z(R)\|_2^2. \tag{29}$$

Here $\mathcal{L}_Z$ is (15), with sample size N. For a PSD matrix the quadratic values are nonnegative, so subtracting their empirical mean is also the minimization over nonnegative common levels used in the Gaussian residual theorem. For $N/d^2 \to \beta > 0$ and standard Gaussian points, the squared-error theorem of Koehler and Sohn [1, Theorems 1.2 and 1.7] gives

$$\Gamma_{N,d}(Z) \xrightarrow{\mathbb{P}} e_\star(\beta, 3), \qquad e_\star(\beta, 3) = \frac{2}{\beta}(-\omega_\beta)_+^2, \qquad v(\omega_\beta) = \beta. \tag{30}$$

Here v is defined in Section 1. It is continuous and strictly decreasing from $+\infty$ to 0 on $(-\infty, \sqrt{2})$, so ω_β is unique and continuous in $\beta > 0$. Thus $e_\star(\beta, 3)$ is continuous, equals zero for $\beta \le 1/4$, and is positive for $\beta > 1/4$, since $v(0) = 1/4$. In particular, for $\beta > 1/4$ there is $\eta_\beta > 0$ such that

$$\mathbb{P}\{\Gamma_{N,d}(Z) \ge \eta_\beta\} \longrightarrow 1. \tag{31}$$

Unlike Lemma 4.1, this input ranges over the entire PSD trace slice. The following consequence is the dual identity of Koehler and Sohn [1, Proposition A.1], in our normalization. We give the construction to make its measurability and sign invariance explicit.

Lemma 6.1 (A certificate determined by the outer products)**.** *Let $N/d^2 \to \beta > 1/4$. There is a Borel choice of $y = y(Z) \in \mathbb{R}^N$, depending only on $(Z_i Z_i^{\mathsf{T}})_{i \le N}$, such that $\sum_i y_i = 0$, $\|y\|_2 \le 1$, and, for any fixed $0 < c_\beta < \sqrt{\beta e_\star(\beta, 3)}$, with probability tending to one,*

$$M_0 := \sum_{i=1}^N y_i Z_i Z_i^{\mathsf{T}} \succeq c_\beta \sqrt{d}\, I_d. \tag{32}$$

Its linear moment $b = \sum_i y_i Z_i$ satisfies $\|b\| = O_{\mathbb{P}}(\sqrt{d})$.

Proof. The image

$$\mathcal{V}_Z = \{\mathcal{L}_Z(R) : R \succeq 0,\ \operatorname{tr} R = d\} \subset \mathbf{1}^\perp$$

is compact and convex. Let $z_*(Z)$ be its unique point of smallest Euclidean norm. Set $y = z_*/\|z_*\|$ when $z_* \ne 0$ and $y = 0$ otherwise. The vector z_* depends continuously on the sample: along a convergent sequence of samples, compactness of the fixed PSD trace slice gives subsequential limits of minimizing matrices, and every resulting image limit is the unique minimizing image at the limiting sample. Consequently y is Borel. Its definition uses only the outer products.

The projection inequality gives, for every $R \succeq 0$ of trace d,

$$\langle y, \mathcal{L}_Z(R) \rangle \ge \|z_*\| = \sqrt{\Gamma_{N,d}(Z)}.$$

This also holds when $z_* = 0$. Since $y \perp \mathbf{1}$, it implies

$$\operatorname{tr}(M_0 R) \ge \sqrt{Nd}\, \sqrt{\Gamma_{N,d}(Z)}.$$

Minimizing the left side over the PSD trace slice gives

$$\lambda_{\min}(M_0) \ge \sqrt{N/d}\, \sqrt{\Gamma_{N,d}(Z)}. \tag{33}$$

Equation (30) proves (32) with every constant c_β in the stated range.

To bound b, represent $Z_i = \varepsilon_i U_i$ with independent standard Gaussian U_i and independent Rademacher signs ε_i. The weights y_i depend on the $U_i U_i^\mathsf{T}$ and hence not on these signs. Conditional on $(U_i)_i$,

$$\mathbb{E}_\varepsilon \left\| \sum_i y_i \varepsilon_i U_i \right\|^2 = \sum_i y_i^2 \|U_i\|^2 \le \max_i \|U_i\|^2.$$

The maximum is at most Cd with probability tending to one because $N = O(d^2)$. Conditional Markov's inequality now gives $\|b\| = O_\mathbb{P}(\sqrt{d})$. □

6.3 Correcting the linear moment on an independent sample

Lemma 6.2 (Correction with a small quadratic contribution)**.** *Suppose $k = k(d)$ satisfies $k/d^2 \to \gamma > 0$. Let H be a $k \times d$ matrix with independent standard Gaussian rows H_j^T, and let $b \in \mathbb{R}^d$ be independent of H, with $\|b\| = O_\mathbb{P}(\sqrt{d})$. Put*

$$P_k = I_k - \frac{1}{k}\mathbf{1}\mathbf{1}^\mathsf{T}, \qquad K = H^\mathsf{T} P_k H, \qquad v = K^{-1} b, \qquad r = P_k H v.$$

For all sufficiently large d, K is invertible almost surely. This supporting assertion is informal and is not formalized; see Appendix A. Define $r = 0$ on the null event of singularity. Then r is Borel, and almost surely

$$\mathbf{1}^\mathsf{T} r = 0, \qquad H^\mathsf{T} r = b. \tag{34}$$

Moreover,

$$\|r\|_2 = O_\mathbb{P}(d^{-1/2}), \qquad \left\| \sum_{j=1}^k r_j H_j H_j^\mathsf{T} \right\|_{\mathrm{op}} = O_\mathbb{P}(\sqrt{\log(2d)}) = o_\mathbb{P}(\sqrt{d}). \tag{35}$$

Proof. The identities in (34) follow directly from the definition. The matrix K has the law of the Gram matrix of $k-1$ independent Gaussian rows. Since $k - 1 \ge d$ eventually, it is invertible almost surely. This step is informal, and the formal proof does not use it. That proof shows instead that the event on which K is singular or the correction is large has probability tending to zero: this is `reserve_correction_converges` in `library/random_matrix/reserve_convergence.v`. On the event $\|k^{-1}K - I_d\|_{\mathrm{op}} \le 1/2$, the lemma `centered_correction_bounds` gives the invertibility of K and (34). Write

$$\overline{H} = \frac{1}{k}\sum_j H_j, \qquad B_j = H_j H_j^\mathsf{T} - I_d, \qquad \mathcal{B}^*(t) = \sum_j t_j B_j.$$

Gaussian covariance concentration and the Gaussian norm bounds give

$$\begin{aligned} \|k^{-1}K - I_d\|_{\mathrm{op}} &= O_\mathbb{P}(d^{-1/2}), & \|H\|_{\mathrm{op}} &= O_\mathbb{P}(d), \\ \|\overline{H}\| &= O_\mathbb{P}(d^{-1/2}), & \|\textstyle\sum_j B_j\|_{\mathrm{op}} &= O_\mathbb{P}(d^{3/2}). \end{aligned} \tag{36}$$

It follows that

$$\|v\| = O_\mathbb{P}(d^{-3/2}), \qquad \left\| v - \frac{b}{k} \right\| = O_\mathbb{P}(d^{-2}), \qquad \|r\|_2^2 = b^\mathsf{T} K^{-1} b = O_\mathbb{P}(d^{-1}). \tag{37}$$

We also use the quadratic-feature bound

$$\|\mathcal{B}^*(t)\|_F \le Ld\,\|t\|_2 \qquad \text{for every } t \in \mathbb{R}^k, \tag{38}$$

on events of probability tending to one, for a fixed L depending on γ. This is the Gaussian case of `mixture_adjoint_bound`, with the factor $\sqrt{d}$ restored because B_j is unnormalized; its analytic source is Koehler and Sohn [1, Proposition 3.1].

Let $D = \sum_j r_j H_j H_j^{\mathsf{T}}$. Since $\sum_j r_j = 0$, equations (36)–(38) give

$$D = D_0 + \mathcal{B}^*\left(H\left(v - \frac{b}{k}\right)\right) - (\overline{H}^{\mathsf{T}} v)\sum_j B_j, \tag{39}$$

$$D_0 = \frac{1}{k}\sum_j (H_j^{\mathsf{T}} b) B_j.$$

The middle term has operator norm at most $Ld\,\|H\|_{\mathrm{op}}\|v - b/k\| = O_{\mathbb{P}}(1)$. The last term has operator norm $O_{\mathbb{P}}(d^{-1/2})$. Thus it suffices to bound D_0.

Represent $H_j = \varepsilon_j U_j$, with independent standard Gaussian U_j and independent Rademacher signs, all independent of b. We may use an event E_d depending only on $(U_j U_j^{\mathsf{T}})_j$, with $\mathbb{P}(E_d) \to 1$, on which (38) and $\max_j \|U_j\|^2 \le L_0 d$ hold for fixed constants L, L_0. On this event, duality in (38), applied to uu^{T} for $\|u\| = 1$, gives

$$\sum_j (U_j^{\mathsf{T}} u)^4 \le 2\sum_j \left((U_j^{\mathsf{T}} u)^2 - 1\right)^2 + 2k \le 2L^2 d^2 + 2k \le L_1 k. \tag{40}$$

Consequently, for all $b, u \in \mathbb{R}^d$ with $\|u\| = 1$,

$$\sum_j (U_j^{\mathsf{T}} b)^2 (U_j^{\mathsf{T}} u)^2 \le L_1 k \|b\|^2,$$

$$\sum_j (U_j^{\mathsf{T}} b)^2 \le \sqrt{L_1}\, k \|b\|^2,$$

by Cauchy–Schwarz and homogeneity. With $B_j = U_j U_j^{\mathsf{T}} - I_d$,

$$B_j^2 = (\|U_j\|^2 - 2) U_j U_j^{\mathsf{T}} + I_d \preceq L_0 d\, U_j U_j^{\mathsf{T}} + I_d.$$

The preceding estimates therefore yield the matrix variance bound

$$\left\| \sum_j (U_j^{\mathsf{T}} b)^2 B_j^2 \right\|_{\mathrm{op}} \le L_2 dk \|b\|^2. \tag{41}$$

Conditional on b and $(U_j)_j$, the matrix $D_0 = k^{-1}\sum_j \varepsilon_j (U_j^{\mathsf{T}} b) B_j$ is a Rademacher series. The expected-norm bound of Tropp [2, Theorem 4.1.1] and (41) give, on E_d,

$$\mathbb{E}_\varepsilon \|D_0\|_{\mathrm{op}} \le \frac{\sqrt{2\log(2d)}}{k} \left\| \sum_j (U_j^{\mathsf{T}} b)^2 B_j^2 \right\|_{\mathrm{op}}^{1/2} \le L_3 \|b\| \sqrt{\frac{d\log(2d)}{k}}. \tag{42}$$

The factor $\sqrt{2\log(2d)}$ is the constant of Tropp's theorem, which is cited and not formalized. The Rocq proof uses `rademacher_mx_series_variance` in `library/random_matrix/rademacher_series.v`, which gives (42) with $40\sqrt{\log(2d)}$ in place of $\sqrt{2\log(2d)}$. Only the constants L_3 and L_4 change, and the conclusion $\|D_0\|_{\mathrm{op}} = O_{\mathbb{P}}(\sqrt{\log(2d)})$ is the same.

For clarity, no moment assumption on the random vector b is needed beyond tightness of $\|b\|/\sqrt{d}$. For $M, T > 0$, conditional Markov's inequality gives, for all sufficiently large d,

$$\mathbb{P}\{\|D_0\|_{\mathrm{op}} > T\sqrt{\log(2d)}\} \le \mathbb{P}(E_d^c) + \mathbb{P}\{\|b\| > M\sqrt{d}\} + \frac{L_4 M}{T}.$$

First choose M large and then T large. This proves $\|D_0\|_{\mathrm{op}} = O_{\mathbb{P}}(\sqrt{\log(2d)})$. Equation (39) proves (35). All constants above are independent of b and d, and may depend on γ. □

6.4 Nonexistence above $1/4$, uniformly over translations

Suppose $\alpha > 1/4$ and fix $\beta \in (1/4, \alpha)$. Split the sample into its first $N = \lfloor \beta d^2 \rfloor$ points and its remaining $k = n(d) - N$ points. Thus $k/d^2 \to \alpha - \beta > 0$. Apply Lemma 6.1 to the first part. It supplies y, M_0, and $b = \sum_{i \le N} y_i Z_i$, with $\|b\| = O_{\mathbb{P}}(\sqrt{d})$ and $M_0 \succeq c_\beta \sqrt{d}\, I_d$ with probability tending to one. The reserve matrix H, with rows Z_{N+j}^{T}, is independent of b. Lemma 6.2 supplies r with $H^{\mathsf{T}} r = b$, $\sum_j r_j = 0$, and $\|D\|_{\mathrm{op}} = o_{\mathbb{P}}(\sqrt{d})$, where $D = \sum_j r_j Z_{N+j} Z_{N+j}^{\mathsf{T}}$.

Define weights on the whole sample by $\lambda = (y_1, \dots, y_N, -r_1, \dots, -r_k)$. They satisfy

$$\sum_i \lambda_i = 0, \qquad \sum_i \lambda_i Z_i = 0, \qquad M := \sum_i \lambda_i Z_i Z_i^{\mathsf{T}} = M_0 - D \succeq \frac{c_\beta}{2} \sqrt{d}\, I_d \tag{43}$$

with probability tending to one. The weights are Borel functions of the sample, using the choices in the two lemmas. We may take Ω_d to be the Borel event on which K is invertible and the last inequality in (43) holds. The two equalities then hold exactly on Ω_d.

For every $m \in \mathbb{R}^d$, expansion of the outer products gives on this same event

$$\sum_i \lambda_i (Z_i + m)(Z_i + m)^{\mathsf{T}} = M. \tag{44}$$

If a nonzero $Q \succeq 0$ fitted the translated points at level one, then

$$0 = \sum_i \lambda_i = \sum_i \lambda_i (Z_i + m)^{\mathsf{T}} Q (Z_i + m) = \operatorname{tr}(QM) \ge \frac{c_\beta}{2} \sqrt{d}\, \operatorname{tr} Q > 0,$$

a contradiction. A fitting matrix is necessarily nonzero, so this excludes all PSD fits. The certificate and the event do not depend on m; the assertion is therefore simultaneous over every translation, without a union bound or a bound on its norm. More generally, for every nonzero $Q \succeq 0$, every $\ell \in \mathbb{R}^d$, and every $c \in \mathbb{R}$,

$$\sum_i \lambda_i (Z_i^{\mathsf{T}} Q Z_i + 2\ell^{\mathsf{T}} Z_i + c) = \operatorname{tr}(QM) > 0.$$

Thus no quadric of the form (6) contains all the points. This includes ellipsoids with arbitrary centers. Finally, Remark 1.2 transfers both threshold conclusions for centered fits to the sign mixtures and completes Theorem 1.3.

6.5 The fitting-error limit for every deterministic shift

The same certificate gives a quantitative residual bound uniformly over translations. Retaining its full margin, and then letting the reserve fraction decrease after taking the dimension limit, identifies the limiting error exactly.

Theorem 6.3 (The Gaussian fitting-error limit is unchanged)**.** *Let $n(d)/d^2 \to \alpha > 0$ and let Z_i be independent standard Gaussian points. For every deterministic sequence $m_d \in \mathbb{R}^d$, with no norm restriction,*

$$\Gamma_{n,d}(Z + m_d) \xrightarrow{\mathbb{P}} e_\star(\alpha, 3). \tag{45}$$

Moreover, for every $\delta > 0$,

$$\mathbb{P}\left\{ \inf_{m \in \mathbb{R}^d} \Gamma_{n,d}(Z + m) \ge e_\star(\alpha, 3) - \delta \right\} \longrightarrow 1. \tag{46}$$

The infimum over all translations also converges in probability to $e_\star(\alpha, 3)$. Equation (45) holds for the sign mixtures in Remark 1.2 as well.

Proof. First suppose $\alpha > 1/4$ and fix $1/4 < \beta < \alpha$. Use the split and certificate of the preceding subsection. On events of probability tending to one, $\Gamma_{N,d}(Z_1, \dots, Z_N) > 0$, so $\|y\|_2 = 1$ and

$$\|\lambda\|_2^2 = 1 + \|r\|_2^2 = 1 + O_{\mathbb{P}}(d^{-1}), \qquad \lambda_{\min}(M) \geq \sqrt{N/d}\sqrt{\Gamma_{N,d}(Z_1, \dots, Z_N)} - \|D\|_{\mathrm{op}}.$$

For every m and every $R \succeq 0$ with $\operatorname{tr} R = d$, the annihilation identities give

$$\langle \lambda, \mathcal{L}_{Z+m}(R) \rangle = \frac{\operatorname{tr}(RM)}{\sqrt{nd}} \geq \sqrt{\frac{d}{n}} \lambda_{\min}(M).$$

Cauchy–Schwarz and minimization over R therefore give, simultaneously for every m,

$$\sqrt{\Gamma_{n,d}(Z+m)} \geq \frac{\left[\sqrt{N/n}\sqrt{\Gamma_{N,d}(Z_1, \dots, Z_N)} - \sqrt{d/n}\|D\|_{\mathrm{op}}\right]_+}{\|\lambda\|_2}. \tag{47}$$

For this fixed β, the square of the right side converges in probability to $(\beta/\alpha)e_\star(\beta, 3)$, because $\|D\|_{\mathrm{op}} = O_{\mathbb{P}}(\sqrt{\log(2d)}) = o_{\mathbb{P}}(\sqrt{d})$. Given $\delta > 0$, continuity in (30) allows us to choose a fixed $\beta < \alpha$ sufficiently close to α to deduce (46). The choice is made before letting $d \to \infty$; no estimate uniform in a vanishing reserve fraction is needed. For $\alpha \leq 1/4$, the same conclusion follows from $\Gamma \geq 0$ and $e_\star(\alpha, 3) = 0$.

For the upper bound, put $p = d - 1$ and choose a deterministic isometry $J_d : \mathbb{R}^p \to \mathbb{R}^d$ with image orthogonal to m_d; if $m_d = 0$, choose any such isometry. The points $g_i = J_d^{\mathsf{T}} Z_i$ are independent standard Gaussian vectors in $\mathbb{R}^p$. For every $A \succeq 0$ with $\operatorname{tr} A = p$, the matrix $R = (d/p) J_d A J_d^{\mathsf{T}}$ is PSD of trace d, and

$$(Z_i + m_d)^{\mathsf{T}} R (Z_i + m_d) = \frac{d}{p}\, g_i^{\mathsf{T}} A g_i.$$

Taking centered residuals with their respective normalizations yields

$$\Gamma_{n,d}(Z + m_d) \leq \frac{d}{p}\, \Gamma_{n,p}(g). \tag{48}$$

Since $n/p^2 \to \alpha$ and $d/p \to 1$, the right side converges to $e_\star(\alpha, 3)$ by (30). Together with (46), this proves (45) at every $\alpha > 0$, including $\alpha = 1/4$. It does not decide exact fitting at the critical density. The infimum over translations is bounded above by $\Gamma_{n,d}(Z)$, which has the same limit. It is measurable because Γ is continuous in the sample and the infimum may be taken over rational translations. Finally, rowwise sign changes leave every quadratic measurement unchanged, proving the assertion for sign mixtures. □

For a simpler positive lower bound at a fixed split, on Ω_d,

$$\inf_{m \in \mathbb{R}^d} \Gamma_{n,d}(Z + m) \geq \frac{c_\beta^2 d^2}{4n\|\lambda\|_2^2} = \frac{c_\beta^2}{4\alpha} + o_{\mathbb{P}}(1).$$

Theorem 6.3 strengthens this bound to the full Gaussian limit.

7 Symmetric mixtures beyond Gaussian data

We now specify a class for which the same certificate works. Let the law of a point in $\mathbb{R}^d$ be a finite or countable mixture

$$\rho_d = \sum_s w_{d,s} \bigotimes_{j=1}^{d} \mu_{d,s,j}, \qquad w_{d,s} \geq 0, \quad \sum_s w_{d,s} = 1. \tag{49}$$

Assume that every component coordinate has mean zero, variance one, and subgaussian norm at most a fixed K_0, and that, for a fixed $\kappa > 1$,

$$\sum_s w_{d,s} \int t^4 \, \mu_{d,s,j}(dt) = \kappa \qquad (1 \le j \le d). \tag{50}$$

These are the admissibility assumptions of the formal mixture theorem. Assume in addition that ρ_d is centrally symmetric: Z and $-Z$ have the same law. This holds, in particular, if every $\mu_{d,s,j}$ is symmetric. No independence of the coordinates under the mixture itself is required.

Write $\alpha_\star(\kappa)$ for the unshifted mixture threshold. In the notation of Section 1, it is

$$\alpha_\star(\kappa) = v(\omega_\kappa) + \frac{\kappa - 3}{2} m(\omega_\kappa)^2, \qquad \omega_\kappa = \frac{\kappa - 3}{2} m(\omega_\kappa),$$

where the root is unique. The threshold lies in $(0, 1/2)$ for every $\kappa > 1$. Let $e_\star(\beta, \kappa)$ denote the unshifted squared-error limit from the same mixture theorem; it is positive when $\beta > \alpha_\star(\kappa)$.

Theorem 7.1 (Translations of symmetric admissible mixtures)**.** *Let $Z_1, \dots, Z_n$ be independent with law* (49) *under the preceding assumptions, and let $n/d^2 \to \alpha > 0$.*

1. *If $\alpha > \alpha_\star(\kappa)$, then on Borel events of probability tending to one there is no nonzero $Q \succeq 0$, $\ell \in \mathbb{R}^d$, and $c \in \mathbb{R}$ with $Z_i^\mathsf{T} Q Z_i + 2\ell^\mathsf{T} Z_i + c = 0$ for all i. In particular, no translated sample $Z + m$ has a PSD fit, simultaneously for all $m \in \mathbb{R}^d$. There is also a constant $\eta > 0$, depending only on α, κ, such that*

$$\mathbb{P}\left\{ \inf_{m \in \mathbb{R}^d} \Gamma_{n,d}(Z + m) \ge \eta \right\} \longrightarrow 1.$$

2. *If $\alpha < \alpha_\star(\kappa)$ and a deterministic shift m_d is supported on $h_d = o(d)$ coordinates, then $Z + m_d$ has a positive definite fit with probability tending to one. There is no restriction on $\|m_d\|$.*

The centered-fitting conclusions also hold for the sign mixture $w_d \,\mathrm{Law}(Z + m_d) + (1 - w_d)\,\mathrm{Law}(Z - m_d)$, for every $w_d \in [0, 1]$.

Proof. Nonexistence. Fix $\alpha_\star(\kappa) < \beta < \alpha$, and split the sample into $N = \lfloor \beta d^2 \rfloor$ points and $k = n - N$ reserve points. The mixture squared-error theorem gives $\Gamma_{N,d} \to e_\star(\beta, \kappa) > 0$ in probability. The compact convex projection in Lemma 6.1 is deterministic and applies without change. It gives a Borel vector y, determined by the outer products, with $\sum_i y_i = 0$, $\|y\|_2 \le 1$, and

$$M_0 \succeq c_\beta \sqrt{d}\, I_d \quad \text{with probability tending to one,} \qquad 0 < c_\beta < \sqrt{\beta e_\star(\beta, \kappa)}.$$

We verify the remaining probabilistic inputs. Conditional on a mixture label, independence and the uniform scalar subgaussian bound give

$$\mathbb{E} \exp(t u^\mathsf{T} Z) \le \exp(C_{K_0} t^2 \|u\|^2), \qquad \mathbb{E} Z Z^\mathsf{T} = I_d.$$

The bound is uniform in the label, so it also holds after mixing. A fixed net of the unit sphere therefore gives $\max_{i \le n} \|Z_i\|^2 \le Cd$ with probability tending to one. Central symmetry permits the representation $Z_i = \varepsilon_i U_i$, where the U_i have law ρ_d and the signs are independent. The conditional second-moment argument in Lemma 6.1 gives $\|b\| = O_\mathbb{P}(\sqrt{d})$ for $b = \sum_{i \le N} y_i Z_i$.

For the reserve rows, the same linear subgaussian bound gives

$$\|k^{-1} H^\mathsf{T} H - I_d\|_{\mathrm{op}} = O_\mathbb{P}(\sqrt{d/k}), \qquad \|\overline{H}\| = O_\mathbb{P}(\sqrt{d/k}).$$

Indeed, for a fixed unit vector u, the independent variables $(u^{\mathsf{T}}H_j)^2-1$ have uniformly bounded subexponential norms. Bernstein's inequality followed by a union bound over a 1/4-net with at most 9^d points proves the covariance estimate. The mean estimate follows from $\mathbb{E}\|\overline{H}\|^2 = d/k$. Since $K/k = H^{\mathsf{T}}H/k - \overline{H}\overline{H}^{\mathsf{T}}$, all the rates in (36) follow. In particular, K is invertible with probability tending to one. Almost-sure invertibility is not needed and is not asserted for discrete laws.

The finite-mixture adjoint bound `mixture_adjoint_bound` gives (38) with a constant depending on K_0 and $\alpha-\beta$. It also holds for countable mixtures: truncate the component labels at tail mass at most d^{-4} and replace the tail by one admissible component. The finite law has the same uniform subgaussian bound, and coupling the $k = O(d^2)$ labels changes a sample with probability at most $kd^{-4} = o(1)$. The adjoint estimate does not require a prescribed fourth moment, so this replacement is sufficient for that estimate. The squared-error input for countable mixtures comes directly from their formal theorem.

Define $v = K^{-1}b$ and $r = P_kHv$ on the invertibility event and set $v = r = 0$ otherwise. Then (34) holds exactly on that event, and (37) holds in probability. The decomposition (39), the fourth-moment estimate (40) obtained from the adjoint bound, and the variance estimate (41) are deterministic. Central symmetry gives the same independent-sign representation of the reserve rows, so (42) and its conditional Markov argument prove

$$\|r\|_2 = O_{\mathbb{P}}(d^{-1/2}), \qquad \|D\|_{\mathrm{op}} = O_{\mathbb{P}}(\sqrt{\log(2d)}) = o_{\mathbb{P}}(\sqrt{d}).$$

Thus $\lambda = (y, -r)$ has all the properties in (43). The deterministic cancellation and quadric argument prove nonexistence on a Borel event independent of the translation. The bound following Theorem 6.3 gives any fixed $0 < \eta < c_\beta^2/(4\alpha)$, proving the residual assertion.

Existence for shifts on few coordinates. Write $h = h_d$, $p = d-h$, and, after a coordinate permutation, $Z_i = (U_i, g_i)$ and $m_d = (\widetilde{m}_d, 0)$, with $a_d = \|\widetilde{m}_d\|$. The case $h = 0$ is the unshifted theorem. The complementary points g_i remain an admissible mixture in dimension p, with the same K_0, κ, and $n/p^2 \to \alpha$. The unshifted mixture theorem supplies, with probability tending to one, $A \succ 0$ with $\operatorname{tr} A = p$, $\operatorname{cond}(A) \le D$ for a fixed D, and $g_i^{\mathsf{T}} A g_i = q > 0$ independent of i.

Set $t_d = (1+a_d)^{-1}$ and

$$y_i = -\frac{t_d(\|U_i + \widetilde{m}_d\|^2 - h - a_d^2)}{\sqrt{p}}.$$

Conditional component independence and variance one imply $\operatorname{Cov}(U_{ij}^2, U_{i\ell}^2) = 0$ for $j \neq \ell$. Central symmetry makes the covariance between $\|U_i\|^2$ and $\widetilde{m}_d^{\mathsf{T}}U_i$ zero. Hence

$$\mathbb{E}y_i^2 = \frac{t_d^2}{p}\big(h(\kappa-1) + 4a_d^2\big), \qquad \frac{\|y\|_2}{p} = O_{\mathbb{P}}\left(\sqrt{\frac{h+1}{d}}\right) = o_{\mathbb{P}}(1).$$

Uniform conditional Bernstein bounds for the sums of squares, and the linear subgaussian bound, also give

$$\|y\|_\infty = O_{\mathbb{P}}\left(\frac{\sqrt{h\log d} + \log d}{\sqrt{d}}\right), \qquad \frac{(\log p)^4}{\sqrt{p}}\|y\|_\infty = o_{\mathbb{P}}(1).$$

Since $\alpha < \alpha_\star(\kappa) < 1/2$, mixture interpolation applies to (g_i) and gives $\|\Delta\|_{\mathrm{op}} = o_{\mathbb{P}}(1)$. It holds simultaneously for every target, so no independence between (U_i) and (g_i) is required. With probability tending to one, $Q = \operatorname{diag}(t_d I_h, A+\Delta)$ is positive definite and has the common quadratic level

$$q + t_d(h + a_d^2) + \operatorname{tr}\Delta + \sqrt{p}\, b_0.$$

This level is positive because $Q \succ 0$ and each $g_i \neq 0$ on the fitting event. Rescaling proves the existence assertion. Finally, central symmetry permits exactly the sign coupling of Remark 1.2, proving the centered-fitting assertions for every mixture weight. □

For example, mixing the Rademacher product law on $\{\pm 1\}^d$ with the standard Gaussian product law, with fixed Gaussian weight $0 < \theta \leq 1$, gives $\kappa = 1 + 2\theta$. The theorem excludes every translated PSD fit above $\alpha_\star(1 + 2\theta)$ and supplies positive definite fits below it for shifts supported on $o(d)$ coordinates. The case $\kappa = 1$ is excluded: unshifted Rademacher data lie exactly on a sphere at every sample size. Existence for arbitrary dense shifts of these non-Gaussian mixtures is not established by the present argument.

A Formal results and their relation to the proof

The conclusions of this note are seven public propositions of the Rocq development, listed below with the seven theorems that prove them. The formalization is released under the annotated tag `shifted` of the repository, placed on 24 October 2026 on commit `305747af48d4f75e26378888a8111d591145fa52`, the final commit of step 18, the step that merged the extension onto phase 14 under the repaired checker. The statement files were accepted at commit `51c2e14d5c2d75469f8147901bc8f664d33d4b6c` on 16 October 2026 and are frozen: rule 36 of `scripts/check_sources.py` compares their code with that commit. The project definitions that they unfold to are either in frozen files or already protected by earlier checks, or among the 29 that `check/ShiftedDefinitions.v` fixes by characterizations proved by computation. Each theorem has an axiom check file, such as `check/ShiftedConditionAssumptions.v`, that compiles only if the theorem proves its frozen statement, and `check/ShiftedIdentities.v` checks the seven together (rule 37); snapshots print the frozen statements and the characterizations against builds of the commits at which they were written, and identity drivers fix the kernel names that they use (rule 38); the check files and the machinery of the extension are pinned, byte for byte, to the integration commit `cc6fc9d49c73f74e0bccdb7f7a7d5716ce02958a` (rule 39). The modules compiled with Rocq 9.1.1 and passed axiom audits, independent source reviews and recursive kernel checking. The references below point to the modules at the tag `shifted`.

The distinction is between conclusions and proof routes: some supporting claims in the text are stronger than the estimates used in Rocq, and only the weaker estimates are formalized. The text marks two of these claims as not formalized: the almost-sure invertibility of K in Lemma 6.2 and the constant $\sqrt{2\log(2d)}$ in (42). The formal proof needs neither. The paragraphs below specify these two differences and several others. They do not list every estimate of the text that the formal proof replaces by a weaker one.

Public statements. The statement files are at the top of `extension/`, and the theorem files are in `extension/Shifted/`.

- Theorem 1.1: `shifted_condition_statement` in `extension/ShiftedConditionStatement.v`, proved by `shifted_condition_theorem` in `extension/Shifted/ConditionTheorem.v`.
- Theorem 1.3, part 1: `shifted_gaussian_statement` in `extension/ShiftedGaussianStatement.v`, proved by `shifted_gaussian_theorem` in `extension/Shifted/GaussianTheorem.v`.
- Theorem 1.3, part 2: `shifted_quadric_statement` in `extension/ShiftedQuadricStatement.v`, proved by `shifted_quadric_theorem` in `extension/Shifted/QuadricTheorem.v`.

- Theorem 6.3: `shifted_error_statement` in `extension/ShiftedErrorStatement.v`, proved by `shifted_error_theorem` in `extension/Shifted/ErrorTheorem.v`.
- Theorem 7.1, part 1: `shifted_mixture_unsat_statement` in `extension/ShiftedMixtureUnsatStatement.v`, proved by `shifted_mixture_unsat_theorem` in `extension/Shifted/MixtureUnsatTheorem.v`.
- Theorem 7.1, part 2: `shifted_mixture_sat_statement` in `extension/ShiftedMixtureSatStatement.v`, proved by `shifted_mixture_sat_theorem` in `extension/Shifted/MixtureSatTheorem.v`. This statement needs no symmetry.
- Remark 1.2 and the sign-mixture sentences of Theorems 1.3, 6.3 and 7.1: `shifted_sign_statement` in `extension/ShiftedSignStatement.v`, proved by `shifted_sign_theorem` in `extension/Shifted/SignTheorem.v`.

Several statements give, for every $\epsilon > 0$ and every sufficiently large dimension, a Borel exceptional set of probability at most ϵ outside which the conclusion holds. In this way they express events that serve all translations, all quadrics or several bounds at once. A diagonal choice gives the events Ω_d of the text, as in Section 2.

Gaussian conclusions. The paths in this paragraph are under `extension/Shifted/`. Theorem 1.1 is `shifted_condition_statement`, and `shifted_condition_theorem` proves it from `gaussian_shift_condition_threshold` and `gaussian_shift_optimal_condition` in `GaussianCondition.v`. The norm hypothesis is exactly $\|m_d\|/\sqrt{d} \to 0$; no logarithmic restriction remains. The optimum is extended-real valued, and its infimum is not assumed attained. The measurability stated in Section 1 is part of `shifted_condition_statement` and `shifted_gaussian_statement`. In `GaussianCondition.v`, `measurable_gaussian_shift_fit` proves it for the events $\{Z + m \in \mathcal{F}_{n,d}(C)\}$ and `measurable_gaussian_shift_condition_number` for $\mathcal{C}(Z + m)$, for every $m \in \mathbb{R}^d$; the case $m = 0$ gives the fitting events and $\mathcal{C}$ themselves.

For Theorem 1.3, part 1 is `shifted_gaussian_statement` and part 2 is `shifted_quadric_statement`. `gaussian_arbitrary_shift_condition_bounds` in `GaussianBounds.v` chooses both constants before the sample-size and shift sequences. The upper event may depend on that deterministic shift sequence. The lower bound in `ConditionLower.v` holds on one event for all translations. Above $1/4$, `gaussian_quadric_good_events` in `GaussianErrorLower.v` proves `shifted_quadric_statement`. It chooses the exceptional event before all Q, v, c, with $Q \succeq 0$ and $Q \neq 0$. Its linear term is $v^{\mathsf{T}}x$; take $v = 2\ell$ to recover the notation of the text.

Theorem 6.3 is `shifted_error_statement`, proved from `gaussian_shift_error_limit` and `gaussian_translation_error_limit` in `GaussianError.v`. These include $\alpha = 1/4$ and arbitrary deterministic shifts. The latter result concerns the infimum over all translations. For positive sample size and dimension, its measurability is proved in `TranslationError.v`, using open strict sublevel sets of an infimum of continuous functions. The all-dimension extension is `measurable_translation_fitting_error_all` in `GaussianErrorLower.v`. No minimizing translation is selected. The uniform lower tail is `gaussian_translation_error_lower_tail` in `GaussianErrorLower.v`. A zero error limit at $1/4$ is not a proof of exact fitting at that density.

Mixtures and signs. Part 1 of Theorem 7.1 is `shifted_mixture_unsat_statement`, proved from `finite_symmetric_quadric` and `countable_symmetric_quadric` in `SymmetricQuadric.v`. Part 2 is `shifted_mixture_sat_statement`, proved from

`finite_sparse_shift_theorem` and `countable_sparse_shift_theorem` in `SparseShift.v`. The nonexistence results require symmetry of the actual mixture point law, not of each component. Their positive residual margin depends only on α, κ and is chosen before the common subgaussian bound and the families.

The sparse-shift existence results are stronger than the theorem stated in the text: they need no symmetry. For the normalized target used in Section 7, `SparseTargets.v` proves

$$\mathbb{E}y_i^2 \le \frac{8K^4h + 8}{p}.$$

It bounds the squared sum of the centered quadratic and linear parts rather than cancelling their covariance. `SparseTargetBounds.v` uses the sufficient maximum bound $\|y\|_\infty = O_{\mathbb{P}}((\sqrt{h}+1)\log(p+1)/\sqrt{p})$. Consequently the interpolation contribution is of order $(\sqrt{h}+1)(\log p)^5/p$, which tends to zero under $h/p \to 0$. There is no concealed condition such as $h(\log p)^a = o(p)$.

The complementary dimension $p(d) = d - h(d)$ need not be injective or monotone. `SparseInputs.v` uses the varying-dimension transfer in `library/analysis/family_reindexing.v`, allowing different laws at repeated values of p. The construction in `SparseExistence.v` uses separate head and complement marginals and a union bound. It never assumes independence of these two blocks after mixing. The final statements have no auxiliary baseline-family or target-tail hypothesis.

The public statement `shifted_sign_statement` collects the sign results. Its first conjunct, the equality of the joint laws of the outer products, holds over any real type. The finite-sample sign identities are in `SignMixture.v` and `library/probability/symmetric_shift.v`. They cover weights in the closed interval $[0, 1]$, including dimension-dependent weights in the limit theorems. `MixtureSign.v` supplies the finite and countable mixture consequences. They transfer centered fit probabilities and the laws of the optimum and fitting error. They do not transfer the quadric event with arbitrary linear terms, or an infimum over further translations: these are not invariant under separate signs on the data.

Inputs and normalization. `GaussianInputs.v` proves standard-normal admissibility with $K = 2$ and fourth moment three. `GaussianExports.v` specializes the existing condition, interpolation, adjoint and error results. `GaussianRows.v` connects flat samples, row tuples, matrices, coordinate splits and rotations. These are proved law identities, not identifications made by notation.

The extracted gap is `gaussian_condition_residual_gap` in `ConditionGap.v`. Its proof retains the loss in `af_value_residual` and bounds the empirical intercept without an exact-fit hypothesis. The existing frozen condition theorem is unchanged. The curve properties remain those of `extension/Condition/Curve.v`; `extension/ConditionKappa/Compatibility.v` identifies this curve and its inverse with the fourth-moment-three specialization.

The identity `sample_residual_gamma` in `ProjectionCertificate.v` identifies the squared closest-residual norm with the frozen `universality_fitting_error`. In particular, the centered squared loss has factor $1/(nd)$, the trace is d, and the best intercept is the empirical mean of nonnegative quadratic measurements. The restriction to nonnegative intercepts therefore causes no change. `library/convex/image_projection.v` proves continuity of the unique closest image point even when its matrix preimage is not unique.

Reserve correction and proof departures. The deterministic correction and cancellation identities are in `library/matrix/centered_correction.v`,

`library/matrix/centered_quadratic_correction.v` and `library/matrix/quadric_certificate.v`. The reserve estimates are in `library/random_matrix/symmetric_reserve.v` and `library/random_matrix/reserve_convergence.v`; `ReserveApplications.v` discharges their hypotheses for Gaussian and admissible mixture rows. The countable adjoint estimate is proved separately by label truncation in `MixtureAdjoint.v`. Row subgaussianity is not used as a substitute for this quadratic-feature estimate.

The matrix Rademacher estimate used formally is `rademacher_mx_series_variance` in `library/random_matrix/rademacher_series.v`. For $d \times d$ matrices its factor is $40\sqrt{\log(2d)}$ instead of $\sqrt{2\log(2d)}$. The file derives it from the existing Gaussian matrix-series estimate and a comparison between signs and Gaussians, rather than formalizing Tropp's matrix Laplace argument. The constant $\sqrt{2\log(2d)}$ in (42) is therefore informal. The formal constant suffices for the reserve error to be $o_{\mathbb{P}}(\sqrt{d})$. The condition lower bound uses a centered sample-covariance deviation estimate for independent isotropic subgaussian rows, together with the raw quadratic adjoint bound. No separate centered Wishart limit is proved.

Lemma 6.2 states that the Gaussian reserve Gram matrix is invertible almost surely for large dimensions. This stronger supporting assertion is marked informal in the text and has not been formalized. Rocq defines the correction to be zero when the matrix is singular, and `reserve_correction_converges` proves that the event on which the matrix is singular or the correction is large has probability tending to zero. That is enough for every theorem of this note. Likewise, the main fitting-error argument fixes a positive reserve density before taking the dimension limit; it needs no estimate uniform in a reserve fraction tending to zero.

Two estimates of Section 3 are also replaced by weaker ones. The text uses $\max_i |\xi_i| = O_{\mathbb{P}}(\sqrt{\log d})$ for (13). The formal proof uses the event $\max_i |\xi_i| \le \log p$, whose complement has probability at most $2n/p^4$ once $\log p \ge 8$ (`gaussian_coordinate_log_tail` in `GaussianTargetBounds.v`); this still gives $S_d = o_{\mathbb{P}}(1)$ when $a_d = o(\sqrt{d})$. The text also says that the common level is positive because the observations are almost surely nonzero. The formal proof uses instead that every point is nonzero on the Gaussian fitting event (`gaussian_conditioned_point_nonzero` and `gaussian_axis_point_nonzero` in `GaussianExistence.v`), so it needs no almost-sure argument.

Verification scope. Each of the seven public theorems uses exactly the usual seven classical axioms of the ellipsoid development: the three `boolp` axioms `propositional_extensionality`, `functional_extensionality_dep` and `constructive_indefinite_description`; `FunctionalExtensionality.functional_extensionality_dep`; `Epsilon.epsilon_statement`; and `sig_forall_dec` and `sig_not_dec` of `ClassicalDedekindReals`. `Classical_Prop.classic` is not among them. The new general library lemmas use at most the three `boolp` axioms. This is not an axiom-free formalization. Source review checks the mathematical meanings and quantifiers; compilation and kernel checking verify the proofs of those propositions. The seven statements are frozen with the repaired checker (rule 36 of `scripts/check_sources.py`; rules 37 to 39 cover the axiom checks, the snapshots and the pin), and `make check` requires their seven axiom reports, 37 reports in all; at the tag `shifted` it passes in two fresh clones of commit `305747af48d4f75e26378888a8111d591145fa52`. Dense non-Gaussian shift existence and exact fitting at the critical density remain outside these conclusions.

References

[1] Frederic Koehler and Youngtak Sohn. Universality and sharp thresholds for ellipsoid fitting, 2026. URL https://arxiv.org/abs/2608.27372v1. arXiv:2608.27372v1.

[2] Joel A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning*, 8(1–2):1–230, 2015. doi: 10.1561/2200000048. URL https://arxiv.org/abs/1501.01571.