Ellipsoid fitting for mixtures of product laws

Statements and proofs

15 October 2026

Abstract

We state the ellipsoid fitting results proved for finite and countable mixtures of product laws. The principal result concerns exact fits with a prescribed condition number. For every fixed fourth moment $\kappa > 1$, it identifies the sharp threshold $\alpha(C, \kappa)$ for an exact positive definite fit with condition number at most C, and the limit $C_\star(\alpha, \kappa)$ of the optimal condition number of an exact fit below the fitting threshold. The curve is the value of a constrained semicircle variational problem and has an explicit integral formula. Both conclusions persist when the point law is replaced by any law at total-variation distance $o(d^{-2})$ from it. At fourth moment three the curve reduces to the earlier one, whose four endpoint asymptotics are proved. We also state the fitting threshold, the limiting fitting error, and interpolation for mixtures, the paired and latent-sign examples, quantitative lower and upper bounds for coordinate variance tensorization, and a continuous latent law singular to every countable mixture of product laws. The original Gaussian theorems and the correspondence with the formal declarations are included. The proofs develop the covariance deformation, the comparison of a diagonal penalty with a trace penalty, the spectral optimization, and the transfer from Gaussian measurements to exact fits.

Contents

1 Setting and notation **2**

2 Mixtures of product laws and the universal threshold **3**

3 The optimal condition number at every fourth moment **8**

4 Two dependent laws with the mixture conclusions **13**

5 Dependence and tensorization of the latent-sign law **15**

6 A continuous mixture separated from all countable mixtures **17**

7 The original Gaussian conclusions **19**

8 Related literature **19**

9 Proof strategy **22**

10 Proof of the scalar formulas **23**

11 The Gaussian cone calculation **28**

12 From Gaussian features to exact fits **33**

13 Proofs for the dependent laws **42**

A Correspondence with the formal statements **46**

B Provenance of the proofs and scope of the contribution **50**

C The reusable Rocq libraries **55**

References **60**

1 Setting and notation

The main result is Theorem 3.6. For independent points whose law is a finite or countable mixture of product laws with mixture-averaged fourth moment $\kappa > 1$, it gives the sharp threshold $\alpha(C, \kappa)$ for an exact fit with condition number at most C, and, for $0 < \alpha < \alpha_\star(\kappa)$, the limit $C_\star(\alpha, \kappa)$ of the optimal condition number. Theorem 3.3 describes the curve and its inverse, and Theorem 3.8 extends both conclusions to perturbed point laws. Section 2 states the admissible mixtures, the unrestricted threshold $\alpha_\star(\kappa)$, the limiting error, and the interpolation estimate that the main result uses. Section 3 contains the condition-number results, with fourth moment three as a specialization. Sections 4–6 treat dependent examples, Section 7 recalls the original Gaussian theorems, Section 8 discusses the literature, and Section 9 explains the proofs. Appendix A lists the formal declarations. The parameters κ, C, and α are fixed as the dimension grows.

Unless explicitly stated otherwise, the dimension d is a positive integer and a sample is an ordered tuple $X = (x_1, \dots, x_n) \in (\mathbb{R}^d)^n$, equipped with its Borel σ-algebra. A point law P_d gives the sample law $P_d^{\otimes n}$: the points are independent and have the same law. Dependence among the coordinates of one point is permitted when the assumptions below permit it. All logarithms are natural.

The sample size in a limit theorem is an arbitrary sequence $n = n(d) \in \mathbb{N}$. The notation

$$\frac{n(d)}{d^2} \longrightarrow \alpha \tag{1}$$

is always taken as $d \to \infty$. Constants declared to depend only on specified parameters are chosen before the families of laws and the sequence $n(d)$. The dimension threshold in an asymptotic conclusion may depend on those families and on the sequence.

Write $\mathbb{S}_d$ for the real symmetric $d \times d$ matrices. The order $A \preceq B$ means that $B - A$ is positive semidefinite, and $S \succ 0$ means that S is positive definite. The trace is $\operatorname{tr} S$, the identity is I_d, and $\|S\|_{\mathrm{op}}$ is the Euclidean operator norm. For vectors, $\|y\|_2$ and $\|y\|_\infty$ are the Euclidean and maximum norms. For symmetric matrices, $\langle A, B\rangle = \operatorname{tr}(AB)$.

1.1 Exact fitting and the optimal condition number

Definition 1.1 (Fitting events)**.** For a positive dimension d, a sample size $n \geq 0$, and $c \in \mathbb{R}$, let

$$\mathcal{F}_{n,d}^{\mathrm{psd}} = \{X : \text{there exists } S \succeq 0 \text{ with } x_i^{\mathsf{T}} S x_i = 1 \text{ for all } i\}, \tag{2}$$

$$\mathcal{F}_{n,d}^{\mathrm{pd}} = \{X : \text{there exists } S \succ 0 \text{ with } x_i^{\mathsf{T}} S x_i = 1 \text{ for all } i\}, \tag{3}$$

$$\mathcal{F}_{n,d}(c) = \left\{ X : \begin{array}{l} \text{there exist } S \in \mathbb{S}_d \text{ and } \ell > 0 \text{ such that} \\ \ell I_d \preceq S \preceq c\ell I_d, \quad x_i^{\mathsf{T}} S x_i = 1 \text{ for all } i \end{array} \right\}. \tag{4}$$

For $S \succ 0$, write $\operatorname{cond}(S) = \lambda_{\max}(S)/\lambda_{\min}(S)$. Thus $\mathcal{F}_{n,d}(c)$ is the event of a positive definite fit with condition number at most c. There is no trace constraint in these events.

The formal positive semidefinite fitting event uses the level d in place of 1. For $d \geq 1$, the map $S \mapsto dS$ identifies the two events. We use level one throughout the extension statements. The Gaussian trace-constrained conclusion in Section 7 retains both its level d and its trace d.

Definition 1.2 (Optimal data condition number)**.** For $X \in (\mathbb{R}^d)^n$, define the extended-real quantity

$$\mathcal{C}(X) = \inf\{c \in \mathbb{R} : X \in \mathcal{F}_{n,d}(c)\}, \qquad \inf \varnothing = +\infty. \tag{5}$$

For $d \geq 1$, this belongs to $[1, +\infty]$. In particular, $\mathcal{C}(X) = +\infty$ if there is no positive definite fit. The use of an infimum does not presume that a minimizing matrix exists.

1.2 The normalized fitting error

Definition 1.3 (Optimal squared error)**.** For $d \geq 1$ and $n \geq 1$, put

$$\Gamma(X) = \inf_{\substack{S \succeq 0,\ \operatorname{tr} S = d \\ r \geq 0}} \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i^{\mathsf{T}} S x_i}{\sqrt{d}} - r \right)^2 . \tag{6}$$

The scalar r is optimized together with S. For the empty sample, set $\Gamma(X) = 0$, as in the formal convention. This statistic uses a trace constraint and a fitted common level; it differs from minimizing the residual against a fixed level one.

1.3 Probability conventions

Convergence in probability to a finite limit uses the extended absolute difference when the statistic can equal $+\infty$. Statements outside measurable exceptional sets do not assert measurability of the full set of samples having the stated property.

For two probabilities P, Q on the same measurable space, use the normalization

$$\mathrm{TV}(P, Q) = \sup_{A \text{ measurable}} |P(A) - Q(A)| \in [0, 1]. \tag{7}$$

2 Mixtures of product laws and the universal threshold

The coordinates of one point need not be independent. They are independent conditional on a component label, and each observation receives its own independent label. The assumptions below allow the component laws and their weights to change with the dimension.

2.1 The scalar threshold and the limiting error

Definition 2.1 (Semicircle functions and the threshold)**.** Let ν_{sc} be the semicircle probability measure of radius $\sqrt{2}$:

$$\nu_{\mathrm{sc}}(dx) = \frac{1}{\pi} \sqrt{2 - x^2}\, \mathbf{1}_{[-\sqrt{2}, \sqrt{2}]}(x)\, dx.$$

In particular, its second moment is $1/2$. For a real number t, set

$$m(t) = \int (x - t)_+ \, \nu_{\mathrm{sc}}(dx), \qquad v(t) = \int (x - t)_+^2 \, \nu_{\mathrm{sc}}(dx), \qquad a_+ = \max\{a, 0\}.$$

For $\kappa > 1$, define

$$E_\kappa(t) = v(t) + \frac{\kappa - 3}{2} m(t)^2, \qquad \omega_\kappa = \sup \left\{ t \in \mathbb{R} : t \leq \frac{\kappa - 3}{2} m(t) \right\}.$$

The fitting threshold is

$$\alpha_\star(\kappa) = E_\kappa(\omega_\kappa). \tag{8}$$

For $\alpha > 0$, define a second cutoff by

$$\omega_{\alpha,\kappa} = \sup \left\{ t \in \mathbb{R} : t < \sqrt{2}, \quad \alpha \leq E_\kappa(t) \right\}.$$

The limiting squared fitting error is

$$e_\star(\alpha, \kappa) = \frac{2}{\alpha} \left(\frac{\kappa - 3}{2} m(\omega_{\alpha,\kappa}) - \omega_{\alpha,\kappa} \right)_+^2 . \tag{9}$$

The supremum definitions specify the cutoffs without making a choice among roots. The following results show that each cutoff is also the unique root of its scalar equation.

Proposition 2.2 (Scalar characterizations)**.** *For every $\kappa > 1$, the number ω_κ belongs to $[-\sqrt{2}, \sqrt{2})$ and is the unique real solution of*

$$t = \frac{\kappa - 3}{2} m(t).$$

For every $\alpha > 0$, the number $\omega_{\alpha,\kappa}$ is the unique solution below $\sqrt{2}$ of $E_\kappa(t) = \alpha$. Moreover,

$$\alpha_\star(\kappa) > 0, \qquad e_\star(\alpha, \kappa) > 0 \quad \Longleftrightarrow \quad \alpha > \alpha_\star(\kappa).$$

Thus $e_\star(\alpha, \kappa) = 0$ for $0 < \alpha \le \alpha_\star(\kappa)$. At fourth moment three, $\omega_3 = 0$ and $\alpha_\star(3) = 1/4$. The threshold also has the variational representation

$$\alpha_\star(\kappa) = \sup_{\substack{f \ge 0,\ f \in L^2(\nu_{\mathrm{sc}}) \\ \int f \, d\nu_{\mathrm{sc}} = 1}} \frac{\left(\int x f(x) \, \nu_{\mathrm{sc}}(dx)\right)^2}{\int f(x)^2 \, \nu_{\mathrm{sc}}(dx) + (\kappa - 3)/2}. \tag{10}$$

Here f is measurable, and the denominator is strictly positive under the displayed constraints.

2.2 Admissible finite and countable mixtures

Definition 2.3 (Admissible mixture families)**.** Fix $K > 0$. In each dimension $d \ge 1$, let the label set J_d be either $\{1, \dots, N_d\}$ for a finite integer $N_d \ge 1$, or $\mathbb{N}$. Let $w_{d,a} \ge 0$ be weights with

$$\sum_{a \in J_d} w_{d,a} = 1.$$

For every $a \in J_d$ and $1 \le j \le d$, let $\mu_{d,a,j}$ be a Borel probability measure on $\mathbb{R}$. The family is *admissible with constant K* if every such coordinate law satisfies

$$\int z \, \mu_{d,a,j}(dz) = 0, \qquad \int z^2 \, \mu_{d,a,j}(dz) = 1, \qquad \int \exp(z^2/K^2) \, \mu_{d,a,j}(dz) \le 2. \tag{11}$$

The associated point law is

$$Q_d = \sum_{a \in J_d} w_{d,a} \bigotimes_{j=1}^{d} \mu_{d,a,j}. \tag{12}$$

For $\kappa > 1$, it has *common fourth moment κ* if

$$\sum_{a \in J_d} w_{d,a} \int z^4 \, \mu_{d,a,j}(dz) = \kappa \qquad (d \ge 1,\ 1 \le j \le d). \tag{13}$$

The constant K is common to all dimensions, labels, and coordinates. Condition (11) applies even to labels whose weight is zero. The fourth moment constraint is imposed after mixing; the separate components need not have fourth moment κ. There is no bound on the growth of N_d and no decay assumption on countably many weights. Within one point, the same label is used for all coordinates. For a sample, the points are independent and have the common law Q_d.

Remark 2.4 (Componentwise isotropy cannot be replaced by global isotropy). Consider the centered, isotropic, uniformly sub-Gaussian law

$$P_d = \tfrac{1}{2}\mathcal{N}(0, \tfrac{1}{2}I_d) + \tfrac{1}{2}\mathcal{N}(0, \tfrac{3}{2}I_d).$$

Every coordinate has law $\mu = \frac{1}{2}\mathcal{N}(0, 1/2) + \frac{1}{2}\mathcal{N}(0, 3/2)$ and fourth moment $15/4$. Nevertheless, if $n(d)/d \to \infty$, independent points with law P_d admit no positive semidefinite exact fit with probability tending to one.

To see this, group the observations by their latent component labels and write $\widehat{M}_-$ and $\widehat{M}_+$ for the empirical second-moment matrices within the two groups. Both group sizes are asymptotic to $n/2$ in probability. Gaussian sample covariance concentration therefore gives

$$\|\widehat{M}_- - \tfrac{1}{2}I_d\|_{\mathrm{op}} \xrightarrow{\mathbb{P}} 0, \qquad \|\widehat{M}_+ - \tfrac{3}{2}I_d\|_{\mathrm{op}} \xrightarrow{\mathbb{P}} 0.$$

Thus $\widehat{M}_+ - \widehat{M}_- \succ 0$ with probability tending to one. An exact fit $S \succeq 0$ would satisfy

$$\mathrm{tr}(S\widehat{M}_-) = \mathrm{tr}(S\widehat{M}_+) = 1, \qquad \mathrm{tr}\big(S(\widehat{M}_+ - \widehat{M}_-)\big) = 0,$$

which is impossible for nonzero $S \succeq 0$.

In contrast, the product law $\mu^{\otimes d}$ has exactly the same coordinate marginals and satisfies Definition 2.3 as a single component. Theorem 2.5 gives it a positive fitting threshold $\alpha_\star(15/4)$. Hence even the full coordinate marginals, together with global isotropy, do not determine the threshold for arbitrary mixtures. This counterexample is an informal observation; it is not part of the Rocq development.

2.3 Fitting and interpolation

Theorem 2.5 (Finite and countable mixture universality)**.** *Fix $\kappa > 1$, $K > 0$, and $\alpha > 0$. The following statements hold for both finite and countable admissible mixture families with common fourth moment κ.*

(i) Below the threshold. *If $\alpha < \alpha_\star(\kappa)$, there exists a constant $c = c(\alpha, \kappa, K) \geq 1$ such that, for every admissible family (Q_d) and every sequence $n = n(d)$ with $n(d)/d^2 \to \alpha$,*

$$Q_d^{\otimes n(d)}(\mathcal{F}^{\mathrm{pd}}_{n(d),d}) \longrightarrow 1, \tag{14}$$

$$Q_d^{\otimes n(d)}(\mathcal{F}_{n(d),d}(c)) \longrightarrow 1. \tag{15}$$

Both events are Borel measurable for every n and d.

(ii) Above the threshold. *If $\alpha > \alpha_\star(\kappa)$, then $e_\star(\alpha, \kappa) > 0$ and, for every admissible family and every such sequence n,*

$$Q_d^{\otimes n(d)}(\mathcal{F}^{\mathrm{psd}}_{n(d),d}) \longrightarrow 0. \tag{16}$$

The fit event is Borel measurable for every n and d.

(iii) The error at every positive density. *For every admissible family, every sequence n with $n(d)/d^2 \to \alpha$, and every $\varepsilon > 0$,*

$$Q_d^{\otimes n(d)}\left(\{X : |\Gamma(X) - e_\star(\alpha, \kappa)| > \varepsilon\}\right) \longrightarrow 0. \tag{17}$$

The displayed deviation event is Borel measurable for every n, d and every real ε.

The constant c is chosen before the mixture family and before $n(d)$.

The frozen finite and countable statements quantify their constants separately. A common choice of c is justified by `countable_of_finite` in `extension/Countable/Theorem.v`: it specializes the countable theorem to finite mixtures without changing the constant. The same specialization preserves the interpolation constant L in Theorem 2.7 below.

The error conclusion includes $\alpha = \alpha_\star(\kappa)$. The theorem makes no claim about the limiting probability of an exact fit at that critical density. The constants are uniform over the specified class of mixture families, but no uniform rate of convergence over that class or over the sample-size sequences is asserted.

The interpolation theorem controls a symmetric correction to a matrix. It does not require a common fourth moment. For a fixed sample $X = (x_1, \dots, x_n)$, write

$$W_i = \frac{x_i x_i^\mathsf{T} - I_d}{\sqrt{d}}, \qquad \langle W_i, \Delta \rangle = \frac{x_i^\mathsf{T} \Delta x_i - \operatorname{tr} \Delta}{\sqrt{d}}.$$

Definition 2.6 (Interpolation property)**.** For $L > 0$, say that X has the interpolation property with constant L if, for every $y = (y_1, \dots, y_n) \in \mathbb{R}^n$, there exist a symmetric matrix $\Delta \in \mathbb{S}_d$ and a real number b such that

$$\langle W_i, \Delta \rangle = y_i + b \qquad (1 \le i \le n) \tag{18}$$

and

$$\|\Delta\|_{\mathrm{op}} \le L \left(\frac{\|y\|_2}{d} + \frac{(\log d)^4}{\sqrt{d}} \|y\|_\infty \right). \tag{19}$$

The trace of Δ is unrestricted, and Δ need not be positive semidefinite. The scalar b is unrestricted as well.

Theorem 2.7 (Interpolation for mixtures)**.** *For every $K > 0$ and $0 < \alpha < 1/2$, there exists $L = L(\alpha, K) > 0$ with the following property. For every finite or countable admissible family (Q_d) with constant K, every sequence n with $n(d)/d^2 \to \alpha$, and every $\varepsilon > 0$, there exists D such that, for every $d \ge D$, there is a Borel set $B_d \subseteq (\mathbb{R}^d)^{n(d)}$ satisfying*

$$Q_d^{\otimes n(d)}(B_d) \le \varepsilon,$$

and every sample outside B_d has the interpolation property with constant L.

The measurable exceptional set is part of the conclusion. The statement does not separately assert measurability of the set of all samples that have the interpolation property.

Corollary 2.8 (Independent coordinates)**.** *Suppose $Q_d = \bigotimes_{j=1}^d \mu_{d,j}$, where every $\mu_{d,j}$ has mean zero, second moment one, fourth moment $\kappa > 1$, and satisfies $\int \exp(z^2/K^2)\, \mu_{d,j}(dz) \le 2$ for the same $K > 0$. For every $\alpha > 0$ and every $n(d)/d^2 \to \alpha$, the following conclusions hold:*

(i) *If $\alpha < \alpha_\star(\kappa)$, the probabilities of a positive definite fit and of a fit with condition number at most $c(\alpha, \kappa, K)$ both tend to one.*

(ii) *If $\alpha > \alpha_\star(\kappa)$, the probability of a positive semidefinite fit tends to zero, and $\Gamma(X) \to e_\star(\alpha, \kappa) > 0$ in probability.*

All the measured events are Borel measurable.

Corollary 2.8 is the content of the separately proved product-law universality statement. Applying Theorem 2.5 to a single component additionally gives the error limit for every $\alpha > 0$, and Theorem 2.7 gives interpolation for $0 < \alpha < 1/2$.

2.4 Exact sample identities and total variation

Proposition 2.9 (Sample laws and their total variation bound)**.** *Let $(Q_a)_{a\in\mathbb{N}}$ be probability measures on a measurable space $\mathcal{X}$, let $w_a \geq 0$ satisfy $\sum_a w_a = 1$, and write $Q = \sum_a w_a Q_a$. For every finite n and every measurable $A \subseteq \mathcal{X}^n$,*

$$Q^{\otimes n}(A) = \sum_{(a_1,\dots,a_n)\in\mathbb{N}^n} \left(\prod_{i=1}^{n} w_{a_i}\right) (Q_{a_1} \otimes \cdots \otimes Q_{a_n})(A). \tag{20}$$

The sum has nonnegative terms and is independent of the enumeration. In particular, substituting $Q_a = \bigotimes_{j=1}^{d} \mu_{d,a,j}$ gives the exact law of the sample of a countable mixture, after identifying the array of nd coordinates with n points of $\mathbb{R}^d$.

For probability measures P, Q on the same space and every measurable $A \subseteq \mathcal{X}^n$,

$$|P^{\otimes n}(A) - Q^{\otimes n}(A)| \leq n \ \mathrm{TV}(P, Q). \tag{21}$$

Consequently, the same bound applies when Q is the countable point mixture in (12)*, without an additional truncation error.*

Theorem 2.10 (Robustness of universality and interpolation)**.** *Fix $\kappa > 1$, $K > 0$, and $\alpha > 0$. Let (Q_d) be a finite or countable admissible mixture family with common fourth moment κ. Let P_d be arbitrary Borel probability measures on $\mathbb{R}^d$ such that*

$$d^2 \ \mathrm{TV}(P_d, Q_d) \longrightarrow 0. \tag{22}$$

For every sequence $n(d)/d^2 \to \alpha$, all three conclusions of Theorem 2.5 hold with $P_d^{\otimes n(d)}$ in place of $Q_d^{\otimes n(d)}$, with a constant $c = c(\alpha, \kappa, K)$ chosen before both (Q_d) and (P_d) and before the sequence $n(d)$.

For interpolation, the fourth moment assumption can be omitted. For every $K > 0$ and $0 < \alpha < 1/2$, there exists $L = L(\alpha, K) > 0$ such that, under $P_d^{\otimes n(d)}$, for every admissible reference family, every (P_d) satisfying (22)*, every sequence $n(d)/d^2 \to \alpha$, and every $\varepsilon > 0$, there exist D and, for each $d \geq D$, a Borel exceptional set of probability at most ε outside which* (18)–(19) *hold for all $y \in \mathbb{R}^{n(d)}$.*

No assumptions on moments, tails, or independence among coordinates are imposed on the perturbed point laws P_d.

The frozen robustness statements have their own existential constants c and L. Reusing the constants from the unperturbed countable theorems is a property of the proofs `robust_statement_holds` and `robust_interpolation_holds` in `extension/Countable/Robust.v`, not an additional clause of those statements. The transfer keeps the conditioned-fit event and the interpolation bound unchanged; only their probabilities and the eventual dimension threshold need to be controlled.

Proposition 2.11 (Contamination at the boundary scale)**.** *Fix $\kappa > 1$, $K > 0$, $0 < \alpha < \alpha_\star(\kappa)$, and $\lambda > 0$. Let (Q_d) be any finite admissible mixture family with common fourth moment κ, and let $n(d)/d^2 \to \alpha$. For every dimension with $n(d) \geq \lambda$, define*

$$P_d^{(\lambda)} = \left(1 - \frac{\lambda}{n(d)}\right) Q_d + \frac{\lambda}{n(d)} \delta_0,$$

where δ_0 is the Dirac probability measure at the zero vector in $\mathbb{R}^d$. At the finitely many remaining dimensions choose any probability law. Whenever $n(d) \geq \lambda$,

$$n(d) \ \mathrm{TV}(P_d^{(\lambda)}, Q_d) \leq \lambda \tag{23}$$

and

$$\big(P_d^{(\lambda)}\big)^{\otimes n(d)}\big(\mathcal{F}^{\mathrm{pd}}_{n(d),d}\big) = \left(1-\frac{\lambda}{n(d)}\right)^{n(d)} Q_d^{\otimes n(d)}\big(\mathcal{F}^{\mathrm{pd}}_{n(d),d}\big), \tag{24}$$

$$\big(P_d^{(\lambda)}\big)^{\otimes n(d)}\big(\mathcal{F}^{\mathrm{pd}}_{n(d),d}\big) \longrightarrow e^{-\lambda} < 1. \tag{25}$$

Identity (24) is the separate lemma `contaminated_probability` in `extension/Countable/Robust.v`. It is not part of the frozen `contamination_statement` in `extension/RobustStatement.v`, which asserts the total variation bound (23) and the limiting probability (25).

This example shows that an $O(d^{-2})$ total variation assumption cannot replace the $o(d^{-2})$ assumption in a theorem guaranteeing fitting probability tending to one for every perturbation. The bound in (23) is an upper bound; equality is not asserted when Q_d has an atom at zero.

3 The optimal condition number at every fourth moment

The principal result identifies, for every fixed fourth moment $\kappa > 1$, the sharp threshold for an exact fit with a prescribed condition number, and the limit of the optimal condition number of an exact fit, an infimum that need not be attained. The fourth-moment assumption is the mixture-averaged condition (13) of Definition 2.3; the components themselves need not have fourth moment κ. The threshold is an explicit curve $\alpha(C,\kappa)$ computed from the semicircle law, and the limiting optimal condition number is its inverse. Section 3.4 records the specialization $\kappa = 3$, where the curve has a simpler description and four endpoint asymptotics are proved.

3.1 The curve and its inverse

We use the semicircle law ν_{sc} of Definition 2.1, of second moment $1/2$, its upper moments m and v, and

$$J(w) = v(w) + w\,m(w) = \int_{\mathbb{R}} z\,(z-w)_+\,\nu_{\mathrm{sc}}(dz), \qquad w \in \mathbb{R}.$$

For $\kappa > 1$, put

$$c_\kappa = \frac{\kappa-3}{2} > -1, \qquad r_\kappa = \frac{c_\kappa}{1+c_\kappa} = \frac{\kappa-3}{\kappa-1} < 1, \qquad a_\kappa = \min\{1, 1+c_\kappa\} > 0.$$

Thus c_κ is half the excess of the fourth moment over its Gaussian value 3; it vanishes exactly at $\kappa = 3$.

Definition 3.1 (The condition-number curve)**.** Let $\kappa > 1$ and $C > 1$. For $u \in \mathbb{R}$, set

$$L_{C,\kappa}(u) = \frac{u + r_\kappa (C-1)^2\, m(u)}{C}, \qquad F_{C,\kappa}(u) = m\big(-L_{C,\kappa}(u)\big) - C\,m(u).$$

Define the root and the lower cutoff by

$$u_{C,\kappa} = \sup\big\{u \in [0,\sqrt{2}] : F_{C,\kappa}(u) \le 0\big\}, \qquad \ell_{C,\kappa} = L_{C,\kappa}(u_{C,\kappa}), \tag{26}$$

and the curve by

$$\alpha(C,\kappa) = J(\ell_{C,\kappa}) - J(u_{C,\kappa}). \tag{27}$$

For $0 < \alpha < \alpha_\star(\kappa)$, define

$$C_\star(\alpha,\kappa) = \inf\{C > 1 : \alpha \le \alpha(C,\kappa)\}. \tag{28}$$

Theorem 3.3 makes these definitions unambiguous on their stated domains.

Definition 3.2 (The constrained variational problem)**.** For $\kappa > 1$ and a Borel function $q : \mathbb{R} \to \mathbb{R}$ with $\int q^2 \, d\nu_{\mathrm{sc}} < \infty$, set

$$\mathcal{V}_\kappa(q) = 2 \int_{\mathbb{R}} z \, q(z) \, \nu_{\mathrm{sc}}(dz) - \int_{\mathbb{R}} q(z)^2 \, \nu_{\mathrm{sc}}(dz) - c_\kappa \left(\int_{\mathbb{R}} q(z) \, \nu_{\mathrm{sc}}(dz) \right)^2 .$$

For $C > 1$, a profile q is *admissible for* C if it is Borel measurable and some $t \geq 0$ satisfies $t \leq q(z) \leq Ct$ for every real z. The zero profile is admissible, and admissible profiles are bounded. The value of the problem is

$$V(C, \kappa) = \sup\{\mathcal{V}_\kappa(q) : q \text{ admissible for } C\}.$$

Theorem 3.3 (The curve and its inverse)**.** *Fix $\kappa > 1$.*

(i) The root. *For every $C > 1$,*

$$0 < u_{C,\kappa} < \sqrt{2}, \qquad -\sqrt{2} < \ell_{C,\kappa} < u_{C,\kappa}, \qquad F_{C,\kappa}(u_{C,\kappa}) = 0,$$

and $u_{C,\kappa}$ is the only zero of $F_{C,\kappa}$ in $[0, \sqrt{2}]$.

(ii) The variational value. *For every $C > 1$, $\alpha(C, \kappa) = V(C, \kappa)$.*

(iii) The integral formula. *For every $C > 1$,*

$$\alpha(C, \kappa) = \frac{1}{3\pi} \int_{\ell_{C,\kappa}}^{u_{C,\kappa}} (2 - z^2)^{3/2} \, dz. \tag{29}$$

(iv) The range. *For every $C > 1$, $0 < \alpha(C, \kappa) < \alpha_\star(\kappa)$. The map $C \mapsto \alpha(C, \kappa)$ is strictly increasing and continuous on $(1, \infty)$, and*

$$\lim_{C \downarrow 1} \alpha(C, \kappa) = 0, \qquad \lim_{C \to \infty} \alpha(C, \kappa) = \alpha_\star(\kappa).$$

(v) The inverse. *For every $0 < \alpha < \alpha_\star(\kappa)$,*

$$C_\star(\alpha, \kappa) > 1, \qquad \alpha(C_\star(\alpha, \kappa), \kappa) = \alpha,$$

and, for every $C > 1$, $\alpha < \alpha(C, \kappa)$ if and only if $C_\star(\alpha, \kappa) < C$.

Moreover, at $\kappa = 3$, $\alpha(C, 3) = \alpha_C$ for every $C > 1$ and $C_\star(\alpha, 3) = C_\star(\alpha)$ for every $0 < \alpha < 1/4$, where α_C and $C_\star(\alpha)$ are the functions of Definition 3.9.

By parts (iv) and (v), the map $C \mapsto \alpha(C, \kappa)$ is a bijection from $(1, \infty)$ onto $(0, \alpha_\star(\kappa))$, with inverse $\alpha \mapsto C_\star(\alpha, \kappa)$. These are fixed-$\kappa$ statements; no uniformity as $\kappa \downarrow 1$ or endpoint rate is asserted.

Proposition 3.4 (The optimizer, the arcsine formula, and almost-everywhere constraints)**.** *Fix $\kappa > 1$ and $C > 1$, and write $u = u_{C,\kappa}$, $\ell = \ell_{C,\kappa}$ and $c = c_\kappa$.*

(a) The optimal profile. *Put*

$$\mu = \frac{(C-1)\, m(u)}{1 + c}, \qquad t = \frac{u - \ell}{C - 1}, \qquad q_\star(z) = \min\{\max\{z - c\mu, \, t\}, \, Ct\}.$$

Then $t > 0$, $\ell = t + c\mu$, $u = Ct + c\mu$, $m(-\ell) = C\, m(u)$, and $\int q_\star \, d\nu_{\mathrm{sc}} = \mu$. The profile $q_\star$ is admissible for C, $\mathcal{V}_\kappa(q_\star) = \alpha(C, \kappa)$, and $\mathcal{V}_\kappa(q) \leq \alpha(C, \kappa)$ for every profile q admissible for C. In particular, the supremum $V(C, \kappa)$ is attained.

(b) The arcsine formula.

$$\alpha(C,\kappa) = A(u) - A(\ell), \qquad A(z) = \frac{\arcsin(z/\sqrt{2})}{2\pi} + \frac{z(5-z^2)\sqrt{2-z^2}}{12\pi}.$$

(c) Almost-everywhere constraints. *The supremum of $\mathcal{V}_\kappa(q)$ over the Borel functions q for which some $t \geq 0$ satisfies $t \leq q \leq Ct$ ν_{sc}-almost everywhere is again $V(C,\kappa) = \alpha(C,\kappa)$.*

(d) Two bounds. *$\alpha(C,\kappa) \leq (C-1)^2/(1+c)$ and $\alpha(C,\kappa) < 1/2$.*

Proposition 3.4 consists of supporting proved lemmas, rather than conjuncts of the frozen curve theorem. Its almost-everywhere equivalence was added after the phase 11 tag. Appendix A gives the precise sources; the proofs are in Section 10.

Remark 3.5 (The range of the optimum). By Proposition 2.2, the cutoff ω_κ satisfies $\omega_\kappa = c_\kappa\, m(\omega_\kappa)$, so

$$\alpha_\star(\kappa) = v(\omega_\kappa) + c_\kappa\, m(\omega_\kappa)^2 = J(\omega_\kappa).$$

The derivative $J'(w) = -(2-w^2)_+^{3/2}/(3\pi)$, which gives (29), shows that J decreases strictly on $[-\sqrt{2},\sqrt{2}]$, with $J(0) = v(0) = 1/4$ and $J(\sqrt{2}) = 0$. Since $m > 0$ on $(-\infty,\sqrt{2})$, the cutoff ω_κ has the sign of c_κ. Hence

$$\alpha_\star(\kappa) = \frac{1}{3\pi}\int_{\omega_\kappa}^{\sqrt{2}} (2-z^2)^{3/2}\,dz,$$

with $\alpha_\star(\kappa) > 1/4$ for $1 < \kappa < 3$ and $\alpha_\star(\kappa) < 1/4$ for $\kappa > 3$. The optimum in Theorem 3.6 is therefore described on $0 < \alpha < \alpha_\star(\kappa)$, an interval that extends beyond $1/4$ when $\kappa < 3$. This is a consequence of the displayed identities, not a separate formal theorem.

3.2 The threshold and the optimum for the data

Theorem 3.6 (Optimal condition number for mixture data)**.** *Fix $\kappa > 1$, $K > 0$, and an admissible finite or countable mixture family $(Q_d)_{d\geq 1}$ as in Definition 2.3, with common fourth moment κ. Let $n = n(d)$ satisfy*

$$\frac{n(d)}{d^2} \longrightarrow \alpha > 0,$$

and let $X^{(d)} = (X_1,\dots,X_{n(d)})$ have law $Q_d^{\otimes n(d)}$. Then the following statements hold.

(i) *For every fixed $C > 1$, the event $\mathcal{F}_{n,d}(C)$ is measurable for every n, d, and*

$$\mathbb{P}\{X^{(d)} \in \mathcal{F}_{n(d),d}(C)\} \longrightarrow \begin{cases} 1, & 0 < \alpha < \alpha(C,\kappa), \\ 0, & \alpha > \alpha(C,\kappa). \end{cases} \tag{30}$$

(ii) *If $0 < \alpha < \alpha_\star(\kappa)$, then*

$$\mathcal{C}(X^{(d)}) \xrightarrow[d\to\infty]{\mathbb{P}} C_\star(\alpha,\kappa). \tag{31}$$

The precise form of this conclusion in the frozen statement is the following: for every $\varepsilon > 0$ and $\delta > 0$ there exists $D \in \mathbb{N}$ such that, for every $d \geq D$, there exists a measurable set $E_d \subseteq (\mathbb{R}^d)^{n(d)}$ with

$$Q_d^{\otimes n(d)}(E_d) \leq \delta, \qquad |\mathcal{C}(x) - C_\star(\alpha,\kappa)| \leq \varepsilon \quad \text{for every } x \notin E_d. \tag{32}$$

In particular, $\mathcal{C}(x)$ is finite on the complement of E_d.

Here $\mathcal{F}_{n,d}(C)$ asks for one positive definite matrix S with $x_i^{\mathsf{T}} S x_i = 1$ for all observations and $\lambda_{\max}(S)/\lambda_{\min}(S) \le C$. The infimum defining $\mathcal{C}$ is over all such exact fits, with value $+\infty$ when there is no positive definite exact fit. The theorem does not require this infimum to be attained.

The parameters κ, K, C, α and the admissible family are fixed before the dimension tends to infinity. The dimension D in (32) may depend on the family, the sequence $n(d)$, ε, and δ. There is no assertion at $\alpha = \alpha(C, \kappa)$ in (30). Above the curve, the threshold covers every $\alpha > \alpha(C, \kappa)$, including $\alpha \ge 1/2$. Nothing is claimed uniformly in κ; the constants of the proof, such as a_κ, deteriorate as $\kappa \downarrow 1$.

Proposition 3.7 (Measurability of the optimal condition number)**.** *For every n, d, the map*

$$x \longmapsto \mathcal{C}(x)$$

is a measurable extended-real-valued function of the sample. In particular, the convergence in (31) *is ordinary convergence in probability of measurable random variables, with an extended real allowed before taking the limit. For every $r \in \mathbb{R}$ its strict sublevel set satisfies*

$$\{x : \mathcal{C}(x) < r\} = \bigcup_{\substack{q \in \mathbb{Q} \\ q < r}} \mathcal{F}_{n,d}(q). \tag{33}$$

This measurability result is a separate proved lemma. The four conjuncts of the frozen condition-number statement are the finite and countable versions of parts (i) and (ii) of Theorem 3.6. Both parts use the explicit curve of Definition 3.1, so the finite and countable versions share the same threshold and the same limit.

3.3 Robustness of the condition-number theorem

The distance between point laws is the total variation distance (7), for which $0 \le \mathrm{TV}(P, Q) \le 1$.

Theorem 3.8 (Robust optimal condition number)**.** *Fix $\kappa > 1$, $K > 0$, and an admissible finite or countable mixture family with common fourth moment κ, whose point laws Q_d are given by* (12)*. Let P_d be any Borel probability law on $\mathbb{R}^d$ such that*

$$d^2 \,\mathrm{TV}(P_d, Q_d) \longrightarrow 0. \tag{34}$$

Let $n(d)/d^2 \to \alpha > 0$ and let $Y^{(d)} = (Y_1, \dots, Y_{n(d)})$ consist of independent points of law P_d. Then:

(i) *For every fixed $C > 1$,*

$$P_d^{\otimes n(d)}\big(\mathcal{F}_{n(d),d}(C)\big) \longrightarrow \begin{cases} 1, & 0 < \alpha < \alpha(C, \kappa), \\ 0, & \alpha > \alpha(C, \kappa). \end{cases} \tag{35}$$

(ii) *If $0 < \alpha < \alpha_\star(\kappa)$, then*

$$\mathcal{C}\big(Y^{(d)}\big) \xrightarrow[d \to \infty]{\mathbb{P}} C_\star(\alpha, \kappa).$$

Explicitly, for every $\varepsilon, \delta > 0$ there exists $D \in \mathbb{N}$ such that for every $d \ge D$ there exists a measurable $E_d \subseteq (\mathbb{R}^d)^{n(d)}$ satisfying

$$P_d^{\otimes n(d)}(E_d) \le \delta, \qquad |\mathcal{C}(x) - C_\star(\alpha, \kappa)| \le \varepsilon \quad \text{for every } x \notin E_d. \tag{36}$$

No moment, tail, coordinate-independence, or mixture representation is assumed for P_d. The fourth moment κ and the component subgaussian bound concern the reference family.

A finite reference family embeds in a countable one by extending the weights by zero and filling inactive labels with admissible laws. This preserves the point law and its total-variation distance. Details are in the proof of the countable and robust results.

3.4 Fourth moment three

At $\kappa = 3$, $c_3 = r_3 = 0$ and $L_{C,3}(u) = u/C$, so $F_{C,3}(Ct) = m(-t) - C\,m(Ct)$. The curve then has the following simpler description, which is the form of the earlier formal statement at fourth moment three.

Definition 3.9 (The condition-number curve at fourth moment three)**.** For $C > 1$, define

$$t_C := \sup\left\{ t \in \left[0, \frac{\sqrt{2}}{C}\right] : m(-t) \le C\,m(Ct) \right\}, \tag{37}$$

$$\alpha_C := \frac{1}{2} - v(-t_C) - v(Ct_C). \tag{38}$$

For $0 < \alpha < 1/4$, define

$$C_\star(\alpha) := \inf\{C > 1 : \alpha \le \alpha_C\}. \tag{39}$$

Proposition 3.10 (Characterization of the curve at fourth moment three)**.** *For every $C > 1$, t_C is the unique solution in $(0, \sqrt{2}/C)$ of*

$$m(-t_C) = C\,m(Ct_C). \tag{40}$$

Moreover $u_{C,3} = Ct_C$ and $\ell_{C,3} = t_C$. The cutoff attains the minimum

$$\alpha_C = \frac{1}{2} - \min_{t\ge 0}\{v(-t) + v(Ct)\} = \frac{1}{2} - \min_{t\ge 0} \int_{\mathbb{R}} \operatorname{dist}(z, [t, Ct])^2\, \nu_{\mathrm{sc}}(dz). \tag{41}$$

The map $C \mapsto \alpha_C$ is continuous and strictly increasing from $(1, \infty)$ onto $(0, 1/4)$. More precisely, for $1 < C \le C'$,

$$0 \le \alpha_{C'} - \alpha_C \le 4(C' - C),$$

and the first inequality is strict when $C < C'$. Its endpoint limits are

$$\lim_{C\downarrow 1} \alpha_C = 0, \qquad \lim_{C\to\infty} \alpha_C = \frac{1}{4}.$$

Consequently, for every $0 < \alpha < 1/4$, $C_\star(\alpha) > 1$ and

$$\alpha_{C_\star(\alpha)} = \alpha. \tag{42}$$

For $C > 1$ the inequalities $C < C_\star(\alpha)$ and $\alpha_C < \alpha$ are equivalent, as are $C > C_\star(\alpha)$ and $\alpha_C > \alpha$.

The compatibility identities follow from the general root equation and identify the earlier fourth-moment-three statements with this specialization. The Lipschitz bound and all four endpoint formulas also have their own formal proofs; see Appendix A.

Theorem 3.11 (The four endpoint asymptotics at fourth moment three)**.** *The functions in Definition 3.9 satisfy the following four statements.*

(i) *There exist $K_1 > 0$ and $u_0 > 0$ such that for every $1 < C < 1 + u_0$,*

$$\left| \alpha_C - \frac{8}{9\pi^2}(C-1)^2(1-(C-1)) \right| \le K_1 (C-1)^4. \tag{43}$$

Equivalently, as $u \downarrow 0$,

$$\alpha_{1+u} = \frac{8}{9\pi^2} u^2 (1-u) + O(u^4).$$

(ii) *For every $\varepsilon > 0$ there exists $C_0 > 1$ such that for every $C \geq C_0$,*

$$\left| C\left(\frac{1}{4} - \alpha_C\right) - \frac{8}{3\pi} \right| \leq \varepsilon. \tag{44}$$

Thus

$$\alpha_C = \frac{1}{4} - \frac{8}{3\pi C} + o(C^{-1}) \qquad (C \to \infty).$$

(iii) *There exist $K_0 > 0$ and $a_0 > 0$ such that for every $0 < \alpha < a_0$,*

$$\left| C_\star(\alpha) - \left(1 + \frac{3\pi}{2\sqrt{2}}\sqrt{\alpha} + \frac{9\pi^2}{16}\alpha\right) \right| \leq K_0 \alpha\sqrt{\alpha}. \tag{45}$$

The neighborhood may be chosen with $a_0 < 1/4$. Equivalently,

$$C_\star(\alpha) = 1 + \frac{3\pi}{2\sqrt{2}}\sqrt{\alpha} + \frac{9\pi^2}{16}\alpha + O(\alpha^{3/2}) \qquad (\alpha \downarrow 0).$$

(iv) *For every $\varepsilon > 0$ there exists $\delta > 0$ such that for every $1/4 - \delta < \alpha < 1/4$,*

$$\left| \left(\frac{1}{4} - \alpha\right) C_\star(\alpha) - \frac{8}{3\pi} \right| \leq \varepsilon. \tag{46}$$

The neighborhood may be chosen with $\delta < 1/4$. Equivalently,

$$C_\star(\alpha) \sim \frac{8}{3\pi(1/4 - \alpha)} \qquad (\alpha \uparrow 1/4).$$

The constants in the remainder bounds are existential. These are asymptotics of the deterministic limiting curve and its inverse at $\kappa = 3$. They apply after taking the dimension limit at fixed C or fixed α in Theorem 3.6; they assert no joint limit in which C or α varies with d. No endpoint expansion is stated or proved for $\kappa \neq 3$; Proposition 3.4 gives only the one-sided bound $\alpha(C, \kappa) \leq (C-1)^2/(1 + c_\kappa)$.

4 Two dependent laws with the mixture conclusions

4.1 The common one-dimensional law

Let μ be the uniform probability measure on $[-\sqrt{3}, \sqrt{3}]$, and put

$$h(t) = \frac{35(t/\sqrt{3})^4 - 30(t/\sqrt{3})^2 + 3}{8}. \tag{47}$$

Thus h is the fourth Legendre polynomial evaluated at $t/\sqrt{3}$. On the support of μ,

$$-\frac{3}{7} \leq h \leq 1, \qquad \int h \, d\mu = 0, \qquad \int h^2 \, d\mu = \frac{1}{9}.$$

The measure μ has mean zero, second moment one and fourth moment $9/5$. For $a \in [-1, 7/3]$, define the probability measure

$$\lambda_a(dt) = (1 + a h(t)) \, \mu(dt). \tag{48}$$

Its first, second and fourth moments are, respectively,

$$0, \qquad 1, \qquad \frac{9}{5} + \frac{8a}{35}.$$

These laws satisfy the common bound $\int \exp(t^2/9) \, \lambda_a(dt) \leq 2$ throughout $a \in [-1, 7/3]$, so the component subgaussian constant $K = 3$ is available uniformly.

4.2 Independent pairs of dependent coordinates

Definition 4.1 (The paired law)**.** For $d \geq 1$ and $-1 \leq \epsilon \leq 7/3$, let $Q_{d,\epsilon}^{\mathrm{pair}}$ be the law on $\mathbb{R}^d$ with density

$$\frac{dQ_{d,\epsilon}^{\mathrm{pair}}}{d\mu^{\otimes d}}(x) = \prod_{k=1}^{\lfloor d/2 \rfloor} (1 + \epsilon h(x_{2k-1})h(x_{2k})). \tag{49}$$

The coordinate pairs are independent. If d is odd, the final coordinate has law μ and is independent of these pairs. Each pair has Lebesgue density

$$\frac{1 + \epsilon h(u)h(v)}{12} \mathbf{1}_{[-\sqrt{3},\sqrt{3}]^2}(u, v).$$

The closed interval $[-1, 7/3]$ is precisely the range in which this density is nonnegative on the square.

Theorem 4.2 (Fitting and interpolation for the paired law)**.** *Fix $\alpha > 0$. Let $\epsilon_d \in [-1, 7/3]$ be any sequence, and let $n = n(d)$ satisfy $n/d^2 \to \alpha$. Write*

$$\mathbb{P}_d^{\mathrm{pair}} = (Q_{d,\epsilon_d}^{\mathrm{pair}})^{\otimes n(d)}$$

for the law of the independent sample $X = (x_1, \ldots, x_n)$. The following statements hold.

(i) If $\alpha < \alpha_\star(9/5)$, there is a constant $c = c(\alpha) \geq 1$ such that

$$\mathbb{P}_d^{\mathrm{pair}}(\mathcal{F}_{n,d}^{\mathrm{pd}}) \longrightarrow 1, \qquad \mathbb{P}_d^{\mathrm{pair}}(\mathcal{F}_{n,d}(c)) \longrightarrow 1.$$

The constant is chosen before the sequences (ϵ_d) and $(n(d))$.

(ii) If $\alpha > \alpha_\star(9/5)$, then $e_\star(\alpha, 9/5) > 0$ and

$$\mathbb{P}_d^{\mathrm{pair}}(\mathcal{F}_{n,d}^{\mathrm{psd}}) \longrightarrow 0.$$

(iii) For every $\alpha > 0$, including the threshold value,

$$\Gamma(X) \xrightarrow[\mathbb{P}_d^{\mathrm{pair}}]{\mathrm{prob.}} e_\star(\alpha, 9/5).$$

Equivalently, for every $\eta > 0$, the probability of $|\Gamma(X) - e_\star(\alpha, 9/5)| > \eta$ tends to zero.

(iv) If $\alpha < 1/2$, there is $C = C(\alpha) > 0$, chosen before (ϵ_d) and $(n(d))$, with the following property. For every $\eta > 0$, for all sufficiently large d, there is a measurable event E_d with $\mathbb{P}_d^{\mathrm{pair}}(E_d) \leq \eta$ such that, for every $X \notin E_d$ and every $y \in \mathbb{R}^n$, there are $\Delta \in \mathbb{S}_d$ and $b \in \mathbb{R}$ satisfying

$$\frac{x_i^\top \Delta x_i - \operatorname{tr} \Delta}{\sqrt{d}} = y_i + b \qquad (1 \leq i \leq n),$$
$$\|\Delta\|_{\mathrm{op}} \leq C \left(\frac{\|y\|_2}{d} + \frac{(\log d)^4}{\sqrt{d}} \|y\|_\infty \right).$$

The fit events and error-deviation events appearing here are measurable for every finite n, d. No conclusion about the fit probability at $\alpha = \alpha_\star(9/5)$ is asserted.

The constants in this theorem are uniform over the whole closed parameter interval, including both endpoints. The theorem permits ϵ_d to vary arbitrarily with d; it does not assert a rate uniform over sample-size sequences.

Formal source: `extension/PairedDefinitions.v`, `paired_sample_density`.

Formal source: `extension/PairedTheorem.v`, `paired_theorem`.

4.3 One latent sign shared by all coordinates

Definition 4.3 (The latent-sign law)**.** For $0 \le s \le 1$, let

$$L_{d,s} = \frac{1}{2}\lambda_s^{\otimes d} + \frac{1}{2}\lambda_{-s}^{\otimes d}, \qquad \rho_{d,s}(x) = \frac{dL_{d,s}}{d\mu^{\otimes d}}(x) = \frac{1}{2}\left\{\prod_{j=1}^{d}(1+sh(x_j)) + \prod_{j=1}^{d}(1-sh(x_j))\right\}. \quad (50)$$

Equivalently, draw one sign with equal probabilities and then, given that sign, draw the coordinates independently from λ_s or λ_{-s}. Different sampled points have independent latent signs.

Theorem 4.4 (Fitting and interpolation for the latent-sign law)**.** *Fix $\alpha > 0$. For any sequence $s_d \in [0,1]$ and any $n(d)$ with $n(d)/d^2 \to \alpha$, let*

$$\mathbb{P}_d^{\text{latent}} = L_{d,s_d}^{\otimes n(d)}.$$

All four conclusions of Theorem 4.2 hold with $\mathbb{P}_d^{\text{latent}}$ in place of $\mathbb{P}_d^{\text{pair}}$:

(i) Below $\alpha_\star(9/5)$, the probabilities of a positive definite exact fit and of an exact fit with condition number at most $c(\alpha) \ge 1$ both tend to one.

(ii) Above $\alpha_\star(9/5)$, the probability of any positive semidefinite exact fit tends to zero, and $e_\star(\alpha, 9/5) > 0$.

(iii) For every $\alpha > 0$, $\Gamma(X) \to e_\star(\alpha, 9/5)$ in probability.

(iv) For $\alpha < 1/2$, the interpolation conclusion in Theorem 4.2(iv) holds outside measurable events of arbitrarily small probability, with a constant $C(\alpha) > 0$.

The constants $c(\alpha)$ and $C(\alpha)$ are chosen before (s_d) and $(n(d))$, and hence work for every such sequence on the closed interval $[0,1]$. The same measurability statements hold. The notation for these constants does not require them to equal the constants chosen for the paired law.

Formal source: `extension/LatentSignDefinitions.v`, `latent_sample_density`.

Formal source: `extension/LatentSignTheorem.v`, `latent_sign_theorem`.

5 Dependence and tensorization of the latent-sign law

5.1 Marginals and covariance identities

Theorem 5.1 (Marginals and fluctuations)**.** *For every $s \in [0,1]$, $d \ge 1$ and $X \sim L_{d,s}$, the following identities hold.*

(i) Each coordinate has law μ, including when $d = 1$. Thus, for every Borel set $A \subseteq \mathbb{R}$ and every $1 \le j \le d$,

$$\mathbb{P}(X_j \in A) = \mu(A), \qquad \mathbb{E}h(X_j) = 0, \qquad \mathbb{E}h(X_j)^2 = \frac{1}{9}.$$

(ii) For distinct coordinates j, k,

$$\operatorname{Cov}(h(X_j), h(X_k)) = \mathbb{E}[h(X_j)h(X_k)] = \frac{s^2}{81}.$$

(iii) With $F_d(x) = d^{-1/2}\sum_{j=1}^{d}|x_j|$,

$$\mathrm{Var}_{L_{d,s}}(F_d) = \frac{1}{4} + \frac{(d-1)s^2}{768}. \tag{51}$$

The all-dimensional marginal assertion uses the separately exported marginal theorem. The frozen scope theorem groups the marginal and centering identities with a second, distinct coordinate, and also asserts (51). The identity for $\mathbb{E}h(X_j)^2$ is proved in the tensorization development.

Formal source: `extension/LatentScopeTheorem.v`, `latent_scope_theorem`.

Formal source: `extension/LatentScopeTheorem.v`, `latent_marginal_theorem`.

5.2 The conditional-variance inequality

Definition 5.2 (Tensorization constant)**.** Write $x^{j\leftarrow t}$ for x with its jth coordinate replaced by t, and define

$$Z_j(x) = \int \rho_{d,s}(x^{j\leftarrow t})\,\mu(dt).$$

For a bounded Borel function $g : \mathbb{R}^d \to \mathbb{R}$, put

$$M_j g(x) = \frac{\int g(x^{j\leftarrow t})\rho_{d,s}(x^{j\leftarrow t})\,\mu(dt)}{Z_j(x)}, \qquad V_j g(x) = M_j(g^2)(x) - (M_j g(x))^2.$$

On $Z_j(x) = 0$ set $M_j g(x) = 0$. These definitions give versions of the conditional expectation and variance given all coordinates other than j, under $L_{d,s}$.

A real number T is a *tensorization constant* for $L_{d,s}$ if, for every globally bounded Borel function $f : \mathbb{R}^d \to \mathbb{R}$,

$$\mathrm{Var}_{L_{d,s}}(f) \le T\sum_{j=1}^{d}\mathbb{E}_{L_{d,s}}[V_j f(X)]. \tag{52}$$

The definition also applies to $d = 0$, with $\mathbb{R}^0$ a singleton and the sum empty; both sides are then zero.

Theorem 5.3 (Lower bounds on every tensorization constant)**.** *Let* $s \in [0,1]$, $d \ge 1$, *and suppose* $T \in \mathbb{R}$ *satisfies* (52)*. Then*

$$T \ge 1 + \frac{(d-1)s^2}{9}. \tag{53}$$

If, in addition,

$$(d-1)s \ge 18, \qquad \exp\left(-\frac{ds^2}{1944}\right) \le \frac{1}{2}, \tag{54}$$

then

$$T \ge \frac{1}{2d}\exp\left(\frac{s^2(d-1)}{15552}\right). \tag{55}$$

Consequently, the constants $c = 1/15552$ *and* $c' = 1/2$*, chosen independently of* s*, satisfy: for every fixed* $s \in (0,1]$ *there is* $D(s) \in \mathbb{N}$ *such that, for every* $d \ge D(s)$ *and every tensorization constant* T*,*

$$T \ge \frac{c'}{d}\exp(cs^2(d-1)).$$

The finite-dimensional exponential estimate is the separately proved lemma underlying the existential constants in the frozen statement. The following auxiliary identities are also proved for $s \in [0,1]$ and $X \sim L_{d,s}$. For $H_d(x) = \sum_{j=1}^{d} h(x_j)$ and $J_d(x) = \mathbf{1}_{\{H_d>0\}}(x)$, with $d \geq 1$,

$$\mathbb{E}H_d(X) = 0, \qquad \mathrm{Var}(H_d(X)) = \frac{d}{9} + \frac{d(d-1)s^2}{81}, \tag{56}$$

$$\frac{1}{4} - \frac{1}{4}\exp\left(-\frac{ds^2}{1944}\right) \leq \mathrm{Var}(J_d(X)) \leq \frac{1}{4}. \tag{57}$$

In particular, $\mathrm{Var}(J_d(X)) \to 1/4$ for each fixed $s > 0$. Although H_d is not globally bounded, its restriction to the support agrees with the globally bounded function $\sum_j h(x_j)\mathbf{1}_{[-\sqrt{3},\sqrt{3}]}(x_j)$ used in the formal tensorization argument.

Formal source: `extension/LatentTensorTheorem.v`, `latent_tensor_theorem`.

Formal source: `extension/LatentTensorTheorem.v`, `latent_tensor_exp`.

Theorem 5.4 (An explicit finite tensorization constant)**.** *For every $s \in [0,1]$ and every integer $d \geq 0$, the law $L_{d,s}$ satisfies* (52) *with*

$$T_0(d,s) = 2\left(\frac{1+s}{1-3s/7}\right)^d. \tag{58}$$

The denominator is strictly positive throughout $[0,1]$. The theorem asserts a valid constant, without asserting its optimality.

For clarity, if $d \geq 1$ we may define in ordinary mathematical notation

$$T_{\mathrm{opt}}(d,s) = \inf\{T \in \mathbb{R} : T \text{ satisfies (52)}\}.$$

The proved bounds give

$$1 + \frac{(d-1)s^2}{9} \leq T_{\mathrm{opt}}(d,s) \leq 2\left(\frac{1+s}{1-3s/7}\right)^d,$$

and, under (54), the exponential expression in (55) is another lower bound. This infimum notation packages the statements about all valid constants; it is not a separate formal definition. No equality of the upper and lower exponential rates is claimed.

Formal source: `extension/LatentTensorUpperTheorem.v`, `latent_tensor_upper_theorem`.

6 A continuous mixture separated from all countable mixtures

Definition 6.1 (The continuous latent law)**.** For $a \in [1,2]$, define the two-point probability measure

$$\nu_a = \frac{1}{1+a^2}\delta_a + \frac{a^2}{1+a^2}\delta_{-1/a}. \tag{59}$$

Let A be uniform on $[1,2]$ and, conditionally on $A = a$, let the d coordinates be independent with law ν_a. The resulting law is

$$P_d^{\mathrm{cont}} = \int_1^2 \nu_a^{\otimes d}\, da. \tag{60}$$

More explicitly, with $v_a(+) = a$, $v_a(-) = -1/a$, $p_a(+) = 1/(1+a^2)$ and $p_a(-) = a^2/(1+a^2)$, it assigns a Borel set $B \subseteq \mathbb{R}^d$ the probability

$$P_d^{\mathrm{cont}}(B) = \int_1^2 \sum_{\sigma \in \{-,+\}^d} \left(\prod_{j=1}^{d} p_a(\sigma_j)\right) \mathbf{1}_B\big(v_a(\sigma_1), \ldots, v_a(\sigma_d)\big)\, da.$$

Theorem 6.2 (Moments and total-variation separation)**.** *The following assertions hold.*

(i) For every $a \in [1,2]$, the component law ν_a has total mass one and

$$\int t\,\nu_a(dt) = 0, \qquad \int t^2\,\nu_a(dt) = 1, \qquad \int t^4\,\nu_a(dt) = a^2 - 1 + a^{-2}.$$

Both of its possible values have absolute value at most 2, and

$$\int_1^2 (a^2 - 1 + a^{-2})\,da = \frac{11}{6}.$$

(ii) The set function in (60) is a Borel probability measure for every $d \geq 0$. For every $d \geq 1$ and every $1 \leq j \leq d$, its coordinate moments are

$$\mathbb{E}_{P_d^{\mathrm{cont}}} X_j = 0, \qquad \mathbb{E}_{P_d^{\mathrm{cont}}} X_j^2 = 1, \qquad \mathbb{E}_{P_d^{\mathrm{cont}}} X_j^4 = \frac{11}{6}.$$

For every nonnegative Borel $g : \mathbb{R} \to [0,\infty)$, its coordinate integration identity is

$$\mathbb{E}_{P_d^{\mathrm{cont}}} g(X_j) = \int_1^2 \left\{ \frac{g(a)}{1+a^2} + \frac{a^2 g(-1/a)}{1+a^2} \right\} da.$$

(iii) Let $d \geq 2$. For any nonnegative weights $(w_r)_{r\geq 0}$ with $\sum_{r\geq 0} w_r = 1$ and any Borel probability measures $\mu_{r,j}$ on $\mathbb{R}$, put

$$Q_d = \sum_{r\geq 0} w_r \bigotimes_{j=1}^{d} \mu_{r,j}.$$

Then

$$\mathrm{TV}(P_d^{\mathrm{cont}}, Q_d) = 1. \tag{61}$$

Here total variation is normalized as $\mathrm{TV}(P,Q) = \sup_{B\ \mathrm{Borel}} |P(B) - Q(B)|$. The weights and product factors are arbitrary: no moments, tail conditions or common marginals are imposed on Q_d.

The separation statement concerns approximation in total variation. It makes no claim that ellipsoid fitting fails for P_d^{cont}. The dimension restriction $d \geq 2$ belongs to the separation assertion; the probability construction and coordinate moments have the wider ranges stated in part (ii).

The frozen theorem contains parts (i) and (iii), including total mass one for $d \geq 2$. The genuine probability construction and coordinate integration and moment identities in part (ii) are separately exported results added in phase 10.

Formal source: `extension/ContinuousLatentTheorem.v`, `continuous_latent_theorem`.

Formal source: `extension/Countable/Continuous.v`, `cl_law`.

Formal source: `extension/Countable/Continuous.v`, `cl_law_mean`.

Formal source: `extension/Countable/Continuous.v`, `cl_law_second_moment`.

Formal source: `extension/Countable/Continuous.v`, `cl_law_fourth_moment`.

7 The original Gaussian conclusions

For completeness, this section states the original three conclusions under the sample law $G_{n,d} = N(0, I_d)^{\otimes n}$. They use the weaker one-sided density hypotheses indicated below, without assuming that $n(d)/d^2$ has a limit.

Theorem 7.1 (Gaussian fitting transition)**.** *The event $\mathcal{F}^{\mathrm{psd}}_{n,d}$ is measurable. For every sequence $n(d)$,*

$$\limsup_{d\to\infty} \frac{n(d)}{d^2} < \frac{1}{4} \quad \Longrightarrow \quad G_{n(d),d}(\mathcal{F}^{\mathrm{psd}}_{n(d),d}) \longrightarrow 1,$$
$$\liminf_{d\to\infty} \frac{n(d)}{d^2} > \frac{1}{4} \quad \Longrightarrow \quad G_{n(d),d}(\mathcal{F}^{\mathrm{psd}}_{n(d),d}) \longrightarrow 0.$$

Theorem 7.2 (A Gaussian fit with prescribed trace)**.** *For every real $\beta < 1/4$, there are constants $0 < \lambda_- \le \lambda_+$ depending only on β such that the event*

$$\mathcal{H}_{n,d}(\lambda_-, \lambda_+) = \left\{ X : \begin{array}{l} \text{there exists } S \in \mathbb{S}_d \text{ with } \operatorname{tr} S = d, \\ \lambda_- I_d \preceq S \preceq \lambda_+ I_d, \quad x_i^{\mathsf{T}} S x_i = d \text{ for all } i \end{array} \right\}$$

is measurable and, for every $n(d)$ satisfying $\limsup_{d\to\infty} n(d)/d^2 \le \beta$,

$$G_{n(d),d}(\mathcal{H}_{n(d),d}(\lambda_-, \lambda_+)) \longrightarrow 1.$$

The constants are chosen before the sequence $n(d)$.

Theorem 7.3 (Balanced positive definite combinations)**.** *The event*

$$\mathcal{B}_{n,d} = \left\{ X : \text{there exists } y \in \mathbb{R}^n, \quad \sum_{i=1}^n y_i = 0, \quad \sum_{i=1}^n y_i x_i x_i^{\mathsf{T}} \succ 0 \right\}$$

is measurable. For every $n(d)$, its probability under $G_{n(d),d}$ tends to zero if $\limsup n(d)/d^2 < 1/4$, and tends to one if $\liminf n(d)/d^2 > 1/4$.

8 Related literature

This section places the preceding statements in the literature, with particular attention to the optimal condition number and the passage from approximate to exact fitting. The references were checked against primary papers and publication records on 9 October 2026. The comparisons for the condition-number results were checked again on 12 October 2026, against the versions listed in the bibliography; the full text of Tanaka and Nakata [17] could not be consulted on that date, and its description below follows the published abstract. The comparison is targeted; it does not establish priority for every extension.

8.1 Ellipsoid fitting and fourth-moment universality

The connection between ellipsoid fitting, correlation matrices, and diagonal plus low-rank matrix decompositions was developed by Saunderson et al. [14]. Saunderson et al. [15] studied Gaussian random instances and presented numerical evidence for the transition at $n \sim d^2/4$. These works explain the connection with minimum-trace factor analysis.

Maillard and Kunisky [9] used the replica method to predict the transition and the spectral shape of fitting ellipsoids. Their predictions include additional spectral constraints and rotationally invariant point laws. Bandeira and Maillard [3] established the sharp threshold

for approximate fitting under spectral restrictions, using a comparison with Gaussian linear measurements. A vanishing average residual leaves open whether every fitting equation can be satisfied exactly.

For Gaussian observations, Misiakiewicz and Wen [10] prove the exact transition, bounded-condition fits below it, and the corresponding alternative theorem. Their results are the source of Section 7. Other 2026 Gaussian threshold results are de la Cerda et al. [6] and Arslan [2]. The Gaussian transition itself is therefore an established theorem in the literature.

The direct antecedent for the extensions is Koehler and Sohn [8, Theorem 1.2]: independent subgaussian coordinates with a common fourth moment have the threshold $\alpha_\star(\kappa)$ and the supercritical error limit $e_\star(\alpha, \kappa)$. Their Theorem 1.7 already gives the error limit for every positive density for centered isotropic data under feature covariance matching, bounded eighth moments, and dimension-free variance tensorization. Remark 1.8 distinguishes this loss statement from exact fitting. The mixture result here covers dependent laws outside those tensorization hypotheses and supplies exact interpolation; the fact that an error limit holds at all positive densities is not by itself a new phenomenon.

8.2 The optimal condition number and its predecessors

The cone of matrices with a prescribed condition bound is an existing object. Tropp [19, Sections 3.1–3.2] defines the cone of positive definite matrices with condition number at most a given constant, proves that it is convex, and computes its dual cone. Bandeira and Maillard [3, Section 3.1.3, Lemma 3.6] study

$$\mathcal{K}_C = \{S \succeq 0 : \lambda_{\max}(S) \le C\lambda_{\min}(S)\}, \qquad C > 1,$$

including zero; the lemma is Lemma 3.5 in the arXiv version. They define an asymptotic Gaussian-width function, prove its monotonicity and its limit as $C \to \infty$, and use a trial spectral profile, the Gaussian spectrum clipped from below, to obtain bounded-condition approximate Gaussian fits. Their calculation does not optimize the full spectral profile. Misiakiewicz and Wen [10, Lemma 2.2] prove a version of the lemma with another trial profile, the positive part of a Gaussian matrix plus a multiple of the identity.

The finite-dimensional projection formula is also known. Tanaka and Nakata [17] reduce nearest positive definite approximation under a condition-number constraint to a simpler problem for unitarily similarity invariant norms, and to a univariate piecewise convex problem for the Ky Fan p-k norms, which include the Frobenius norm. For the Frobenius norm, the reduced problem clips the eigenvalues to $[t, Ct]$ and minimizes over the scalar $t \ge 0$, as in (86). Related clipping and scalar optimization appear in the covariance likelihood problem of Won et al. [20]; its objective differs from Frobenius distance. General Gaussian conic transitions and the relation between Gaussian width and statistical dimension are established in Amelunxen et al. [1, Theorem I and Proposition 10.2].

At $\kappa = 3$, the curve of Definition 3.9 evaluates this spectral optimization for the cone under the semicircle law, and identifies the optimal interval through $m(-t_C) = Cm(Ct_C)$. For $\kappa \neq 3$, the fourth moment deforms the feature covariance. The Gaussian problem then concerns the transformed cone $T_\kappa \mathcal{K}_C$, whose support function is not a function of a spectrum alone, and the curve $\alpha(C, \kappa)$ of Definition 3.1 is the value of a constrained semicircle problem with a trace term. Theorem 3.6 transfers the threshold to the rank-one quadratic measurements of the data, for finite and countable mixtures with averaged fourth moment κ, and identifies the optimum over exact fits. Theorem 3.8 extends both conclusions to nearby point laws.

Existence of some bounded condition number below the fitting threshold was already known. For Gaussian data, Misiakiewicz and Wen [10, Theorem 1.3 and Remark 3.6] give exact fits with a condition bound of order $(1/4 - \alpha)^{-2}$ near the threshold. Their remark states that this bounds the fit produced by the proof, not every fit; the square comes from replacing

the condition cone by a spectral box in that proof. For independent coordinates with fourth moment $\kappa > 1$, Koehler and Sohn [8, Theorem 1.2(i) and Lemma 5.17] show that the fits below $\alpha_\star(\kappa)$ can be chosen with bounded condition number, with constants depending only on κ and on the distance from α to $\alpha_\star(\kappa)$; they do not optimize the ratio. Theorem 2.5 gives the same existence for mixtures.

At $\kappa = 3$, $C_\star(\alpha)$ has order $(1/4 - \alpha)^{-1}$ near the threshold, with coefficient $8/(3\pi)$ (Theorem 3.11). A short calculation, not stated in the cited papers, shows that the trial profiles of Bandeira and Maillard [3] and Misiakiewicz and Wen [10] already give Gaussian-model ratios with the same leading term. For instance, taking $t = \sqrt{2}/C$ in (41), which corresponds to the semicircle spectrum clipped from below only, gives $1/2 - v(-\sqrt{2}/C) = 1/4 - 8/(3\pi C) + O(C^{-2})$, while $\alpha_C = 1/4 - 8/(3\pi C) + o(C^{-1})$. For the positive-part profile $q(z) = z_+ + s$, optimization of its scale gives the squared support value

$$\frac{(\int z q(z)\, d\nu_{\mathrm{sc}})^2}{\int q(z)^2\, d\nu_{\mathrm{sc}}} = \frac{1/16}{1/4 + 2m(0)s + s^2} = \frac{1}{4} - \frac{8}{3\pi C} + O(C^{-2}), \qquad s = \frac{\sqrt{2}}{C - 1}.$$

For the random-matrix trials, a spectral-edge cutoff strictly above $\sqrt{2}$ is used first and then decreased to $\sqrt{2}$ after the dimension limit. The coefficient is therefore already accessible from those constructions. The new elements are the matching lower bound, which excludes fits with a smaller ratio, the exact curve at every C and κ, and the transfer of both to exact fits on data. These comparisons take $d \to \infty$ at fixed α before taking $\alpha \uparrow 1/4$.

The closest antecedent we found is a non-rigorous, one-sided analogue. By the replica method and for Gaussian data, Maillard and Kunisky [9, Section 3.3] predict the threshold for fits with $S \succeq \lambda I_d$ (their κ), and in their Section 1.3.1 they invert it to predict the largest attainable smallest eigenvalue. Since the trace of any fit concentrates [9, Section 1.1], this constraint bounds the smallest eigenvalue relative to the mean eigenvalue. It is one-sided: nothing bounds $\lambda_{\max}(S)$, so it is not a condition bound.

8.3 Universality methods, mixtures, and exact interpolation

The comparison of optimized empirical risks with Gaussian models is a broader theme. Montanari and Saeed [11] prove universality under hypotheses that control Gaussian approximation on the optimization domain. Dandi et al. [5] develop this program for mixtures in generalized linear models: their Theorem 1 and Corollary 2 concern free energies and minimum training risks under compactness, Lipschitz, and conditional central-limit hypotheses. Their asymptotic formulation keeps the component set fixed. These results give a methodological precedent for mixture universality. Applying them to quadratic features would still require checking their hypotheses and producing an exact positive definite interpolant.

The mixture proof here uses two different facts. Conditional on all labels, the interpolation estimates hold uniformly over the product components. Unconditionally, the quadratic features retain the required covariance and variance bounds. Component fourth moments may differ, so applying an entire common-fourth-moment theorem separately within each component would not justify the conclusion. Uniform estimates permit a growing finite number of labels; truncation then gives the countable extension with the same constants.

The Gaussian part uses established comparison principles: Gordon's escape theorem [7], the convex Gaussian min–max method [18], and the conic geometry described by Amelunxen et al. [1]. Small-ball estimates are the anti-concentration input for exact interpolation. The projection estimates of Rudelson and Vershynin [12] are a close antecedent; the product-law interpolation argument here uses a moment-based version that also permits discrete coordinates. Conditioning that estimate on labels is a derived use of independent-variable theory.

8.4 Dependent examples, robustness, and the scope of novelty

The paired density is an instance of the classical Sarmanov construction [13]: bounded centered kernels perturb a product density by a factor $1 + \epsilon h(x)h(y)$. The Legendre kernel, the full interval $[-1, 7/3]$, and the fitting conclusions are the explicit specialization studied here. Its role is to illustrate the mixture theorem.

The deterioration of functional inequalities under mixing is also well established; see Schlichting [16] for Poincaré and logarithmic Sobolev inequalities for two-component mixtures. That work uses gradient energies. Theorems 5.3 and 5.4 concern the sum of coordinate conditional variances instead. For every fixed $s > 0$, their explicit bounds show that exact fitting in the latent-sign example coexists with an exponentially large tensorization constant. The lower bound detects the latent component with a bounded function; the upper bound compares densities with a product law.

The robustness theorems use the elementary product-law estimate (21). Their content is the transfer of the proved fitting and optimal-condition conclusions without imposing moments on the perturbed laws. The contamination example identifies the limiting cost of a rare impossible observation. The continuous latent example is a separate measure-theoretic obstruction: countable product mixtures need not approximate a continuously mixed law in total variation. Its support argument is described in Section 9; it is not a fitting impossibility result.

No equivalent theorem for the full finite/countable mixture class was identified in the sources checked. For the condition number, the sources checked on 12 October 2026, Maillard and Kunisky [9], Bandeira and Maillard [3], Misiakiewicz and Wen [10], de la Cerda et al. [6], Koehler and Sohn [8], and Arslan [2], state no threshold for exact fits with condition number at most C, no explicit curve $\alpha(C)$ or $\alpha(C, \kappa)$, and no limit for the optimal condition number of exact fits, for Gaussian or non-Gaussian data. This is a qualified literature finding from a targeted search, not a claim of exhaustive novelty. The established threshold formulas, spectral clipping, Gaussian comparison tools, and classical density construction are credited above. Formal verification establishes the stated implications from their assumptions; a publication-priority claim requires a separate assessment.

9 Proof strategy

The proof has three parts. First, a scalar variational problem determines $\alpha(C, \kappa)$. Second, this value gives the Gaussian size of the condition cone after the covariance deformation induced by κ. Third, comparison with quadratic data features and an interpolation estimate turn the Gaussian threshold into a threshold for exact fits. Sections 10–12 give these arguments; Section 13 proves the dependent-law examples.

The main difficulty in the Gaussian calculation is that the covariance map T_κ scales the diagonal entries of a matrix, so $T_\kappa \mathcal{K}_C$ is not invariant under orthogonal conjugation when $\kappa \neq 3$. A spectral calculation cannot be applied to this cone directly. We first compare its quadratic optimization problem with one whose penalty depends only on the trace. The latter has a unique conjugation-equivariant maximizer. Rotation invariance then controls its diagonal fluctuations, and a coercive quadratic remainder controls how much the original maximum can exceed its value at this maximizer. Both estimates are needed. They work for every $c_\kappa > -1$, including the negative-penalty regime $1 < \kappa < 3$.

A scalar shift removes the trace penalty. The resulting projection onto $\mathcal{K}_C$ clips eigenvalues to $[t, Ct]$. Uniform semicircle convergence over the shift and level, followed by the scalar dual identity, gives the Gaussian support limit. The direction of the dual optimization changes at $\kappa = 3$; the proof handles both signs explicitly. Gaussian min–max comparison then supplies a bounded residual below the curve and a positive residual gap above it.

For the data, covariance matching is only the first requirement. Componentwise product estimates must also survive averaging over the labels. In particular, the variance of a quadratic form in the feature vector acquires a covariance-spread term, and the interpolation bounds must be uniform over all label tuples. A finite cover of spectral slices lets the regularized comparison retain a strict condition-number margin. Below the curve, a small operator-norm correction removes every residual. Above it, a bound on the free intercept allows the Gaussian gap to exclude an exact fit. Bracketing by two fixed condition bounds then controls the optimal condition number without assuming its infimum is attained.

Countable mixtures follow by replacing the tail with one product law that preserves the coordinate moments. Total-variation robustness uses the exact sample-law bound (21). These two transfers do not change the matrix inequalities or their constants.

The proofs below state the established comparison and interpolation inputs where they are used, with their hypotheses and sources. The scalar optimization, covariance reduction, finite-mixture application, and transfers are proved here. The earlier unrestricted Gaussian error theorem is used as a cited input.

10 Proof of the scalar formulas

Throughout this section, expectations are against $Z \sim \nu_{\mathrm{sc}}$, $b = \sqrt{2}$, $c = (\kappa - 3)/2 > -1$, and $a = \min\{1, 1 + c\} > 0$. We prove the scalar characterizations, the condition-number curve, and its four stated endpoint asymptotics. The argument that identifies this scalar curve with the data threshold is separate.

10.1 The unrestricted threshold and error

Symmetry and differentiation under the integral give

$$m(-w) = m(w) + w, \qquad v(-w) + v(w) = \tfrac{1}{2} + w^2, \tag{62}$$

$$m'(w) = -\mathbb{P}(Z > w), \qquad v'(w) = -2m(w). \tag{63}$$

The differentiations are justified by the bounded support and absence of atoms. In particular, m is 1-Lipschitz, is strictly decreasing on $(-\infty, b)$, is positive there, and vanishes on $[b, \infty)$. For $w \le -b$, $m(w) = -w$ and $v(w) = 1/2 + w^2$.

Write $h(w) = w - cm(w)$. For $x < y$, the Lipschitz bound gives

$$h(y) - h(x) \ge a(y - x).$$

Also $h(-b) = -(1 + c)b < 0 < h(b) = b$. Consequently h has a unique zero $\omega = \omega_\kappa \in (-b, b)$, and its sublevel set is $(-\infty, \omega]$, agreeing with the supremum in Definition 2.1. Similarly,

$$E'_\kappa(w) = -2m(w)(1 + c\mathbb{P}(Z > w)) < 0 \qquad (w < b).$$

Since $E_\kappa(b) = 0$ and $E_\kappa(w) = 1/2 + (1 + c)w^2 \to \infty$ as $w \to -\infty$, every $\alpha > 0$ has exactly one level cutoff $\omega_{\alpha,\kappa} < b$. The sign of $cm(\omega_{\alpha,\kappa}) - \omega_{\alpha,\kappa}$ is positive exactly when $\omega_{\alpha,\kappa} < \omega$, or equivalently $\alpha > \alpha_\star(\kappa)$. This proves the assertion about $e_\star$ in Proposition 2.2.

The variational formula requires some care when $c < 0$. On $L^2(\nu_{\mathrm{sc}})$ define the bilinear form

$$B_c(f, g) = \mathbb{E}[fg] + c\mathbb{E}[f]\mathbb{E}[g].$$

It is an inner product, because

$$B_c(f, f) \ge a\mathbb{E}[f^2]. \tag{64}$$

Put $p(z) = (z-\omega)_+$. Since $\omega = c\mathbb{E}[p]$, the residual $z - \omega - p(z)$ is nonpositive and vanishes wherever $p > 0$. Thus, for every nonnegative $f \in L^2(\nu_{\mathrm{sc}})$,

$$\mathbb{E}[Zf] \le B_c(p,f), \qquad B_c(p,p) = \mathbb{E}[Zp] = v(\omega) + cm(\omega)^2 = \alpha_\star(\kappa).$$

Cauchy–Schwarz for B_c bounds $(\mathbb{E}[Zf])_+^2$ by $\alpha_\star(\kappa)B_c(f,f)$. Applying the same inequality to $f(-z)$ bounds the negative part as well. Dividing by $(\mathbb{E}f)^2$ proves (10); equality holds at $f = p/m(\omega)$. In particular, the threshold is positive.

For later use, the same argument with $Z + z$ in place of Z shows that the unique solution $w + z = cm(w)$ satisfies

$$\sup_{\substack{f\ge 0, f\in L^2(\nu_{\mathrm{sc}})\\ \mathbb{E}f>0}} \frac{(\mathbb{E}[(Z+z)f])_+^2}{B_c(f,f)} = v(w) + cm(w)^2. \tag{65}$$

If this value is positive, it is attained at $(Z-w)_+$; if it is zero, the preceding upper bound proves the identity directly. This also explains the scalar quantity used in the limiting error.

Denote the value in (65) by $\chi_c(z)$. The cutoff is continuously differentiable in z, since $h'(w) = 1 + c\mathbb{P}(Z > w) \ge a$, and the chain rule gives $\chi_c'(z) = 2m(w)$. Thus χ_c is nondecreasing and is strictly increasing wherever it is positive. There is also an exact duality. For $c^\vee = -c/(1+c)$ and $z^\vee = -z/(1+c)$, symmetry shows that the dual cutoff is $-w$. Using (62) and $w + z = cm(w)$ gives

$$\chi_c(z) + \chi_{c^\vee}(z^\vee) = \frac{1}{2} + \frac{z^2}{1+c}. \tag{66}$$

In fourth-moment notation, $\kappa^\vee = 1 + 4/(\kappa - 1)$ and $\alpha_\star(\kappa) + \alpha_\star(\kappa^\vee) = 1/2$.

10.2 Existence and uniqueness of the two clipping points

Fix $C > 1$ and put $r = c/(1+c) < 1$. We abbreviate the functions of Definition 3.1 to L and F. One must not assume that L is increasing. Instead, for $x < y \le b$, put $\Delta = m(x) - m(y) > 0$. If $L(x) \le L(y)$, monotonicity of m gives $F(y) - F(x) \ge C\Delta > 0$. Otherwise its Lipschitz bound gives

$$\begin{aligned} C(F(y) - F(x)) &\ge C^2\Delta - C(L(x) - L(y)) \\ &= (y-x) + (C^2 - r(C-1)^2)\Delta > 0. \end{aligned}$$

Thus F is strictly increasing on $[0,b]$, and it is continuous. At b, $F(b) = m(-b/C) > 0$. At zero, if $L(0) \le 0$ then $m(-L(0)) \le m(0) < Cm(0)$. If $L(0) > 0$, then

$$m(-L(0)) = m(L(0)) + L(0) \le m(0) + L(0) < Cm(0),$$

because $L(0) = r(C-1)^2 m(0)/C < (C-1)m(0)$. There is therefore a unique zero $u \in (0,b)$, and the supremum definition selects exactly this zero.

Set $\ell = L(u)$. The balance equation is

$$m(-\ell) = Cm(u). \tag{67}$$

Its right side is positive, so $\ell > -b$. If $\ell \ge u$, then symmetry and monotonicity give $Cm(u) = m(\ell) + \ell \le m(u) + \ell$, hence

$$C\ell - u \ge (C-1)\ell \ge (C-1)^2 m(u).$$

But the definition of L says that the left side is $r(C-1)^2 m(u)$, a contradiction. Thus $-b < \ell < u < b$.

10.3 Optimality and value

Define

$$\mu = \frac{(C-1)m(u)}{1+c}, \qquad t = \frac{u-\ell}{C-1} > 0, \qquad q_\star(z) = \min\{\max\{z - c\mu, t\}, Ct\}.$$

The identity $C\ell - u = r(C-1)^2 m(u)$ gives $\ell = t + c\mu$ and $u = Ct + c\mu$. Since

$$q_\star(z) = t + (z-\ell)_+ - (z-u)_+, \tag{68}$$

its mean is

$$\mathbb{E}q_\star = t + m(\ell) - m(u) = (C-1)m(u) - c\mu = \mu.$$

Here is the first-order inequality that proves global optimality. The residual is

$$R(z) = z - c\mu - q_\star(z) = (z-u)_+ - (\ell - z)_+.$$

If a competing profile satisfies $s \le q \le Cs$ for some $s \ge 0$, then (67) yields

$$\mathbb{E}[Rq] \le Cs\, m(u) - s\, m(-\ell) = 0, \qquad \mathbb{E}[Rq_\star] = Ct\, m(u) - t\, m(-\ell) = 0.$$

Writing $h = q - q_\star$, expansion of the objective now gives

$$\mathcal{V}_\kappa(q) - \mathcal{V}_\kappa(q_\star) = 2\mathbb{E}[Rh] - B_c(h,h) \le -a\mathbb{E}[h^2]. \tag{69}$$

This proves optimality, including the case $1 < \kappa < 3$, where the mean penalty has the opposite sign. It also proves uniqueness up to ν_{sc}-null sets.

The radial identity $\mathbb{E}[Rq_\star] = 0$ implies

$$\mathcal{V}_\kappa(q_\star) = B_c(q_\star, q_\star) = \mathbb{E}[Zq_\star] = J(\ell) - J(u) = \alpha(C,\kappa). \tag{70}$$

The last equality uses (68) and $\mathbb{E}Z = 0$. For a Borel profile bounded between s and Cs only almost everywhere, replace it by $\min\{\max\{q,s\}, Cs\}$. This changes none of its integrals and makes it admissible everywhere. Hence the two admissible classes have the same supremum, proving Proposition 3.4(c).

10.4 The signed scalar dual

The spectral argument uses an equivalent scalar optimization whose direction depends on the sign of c. For $\beta \in \mathbb{R}$ set

$$D(\beta) = \frac{1}{2} + c(1+c)\beta^2 - \inf_{s \ge 0}\{v(-s - c\beta) + v(Cs + c\beta)\}. \tag{71}$$

Pointwise projection onto $[s, Cs]$, followed by minimization over s, shows that

$$\begin{aligned} D(\beta) &= \sup_{q \text{ admissible for } C} \{2\mathbb{E}[(Z - c\beta)q] - \mathbb{E}[q^2] + c\beta^2\} \\ &= \sup_{q \text{ admissible for } C} \{\mathcal{V}_\kappa(q) + c(\mathbb{E}q - \beta)^2\}. \end{aligned} \tag{72}$$

The term $c^2\beta^2$ in the first line of (71) comes from $\mathbb{E}(Z - c\beta)^2$. At $\beta = \mu$, the same residual inequality as above, now with quadratic term $\mathbb{E}[h^2]$, proves that $q_\star$ maximizes the shifted objective. Thus $D(\mu) = \alpha(C,\kappa)$ and

$$\alpha(C,\kappa) = \begin{cases} \max\limits_{\beta \in \mathbb{R}} D(\beta), & -1 < c < 0, \\ D(0), & c = 0, \\ \min\limits_{\beta \in \mathbb{R}} D(\beta), & c > 0. \end{cases} \tag{73}$$

Indeed, for $c < 0$ each term in the supremum in (72) is at most the optimum; for $c > 0$ testing $q_\star$ gives the reverse inequality. Moreover, (64) and (70) give $a\|q_\star\|_2^2 \le \|q_\star\|_2/\sqrt{2}$, so $0 < \mu \le 1/(\sqrt{2}a)$. The optimizing β therefore lies in the compact interval $[0, 1/a]$ used in the spectral limit.

10.5 Integral formula, range, and inverse

For $-b < w < b$, differentiation of $J(w) = \mathbb{E}[Z(Z-w)_+]$ gives

$$J'(w) = -\int_w^b \frac{z\sqrt{2-z^2}}{\pi}\,dz = -\frac{(2-w^2)^{3/2}}{3\pi}. \tag{74}$$

There is no boundary contribution because $(z-w)_+$ vanishes at $z=w$; dominated convergence justifies the differentiation. Integrating between ℓ and u proves the integral formula. Differentiating the function A in Proposition 3.4 gives $A'(w) = (2-w^2)^{3/2}/(3\pi)$ on this interval, proving the arcsine formula. Also $J(-b) = 1/2$, $J(b) = 0$, and

$$\alpha_\star(\kappa) = v(\omega) + cm(\omega)^2 = J(\omega) \in (0, 1/2).$$

The sign of ω is the sign of c, so this proves the comparisons with $1/4$ in Remark 3.5.

Admissible classes increase with C, but strict increase needs an actual improvement of a profile. For $C' > C$, let $p_u = \mathbb{P}(Z > u) > 0$ and choose $\varepsilon > 0$ sufficiently small that $\varepsilon \le (C'-C)t$. The function $q_\star + \varepsilon \mathbf{1}_{(u,\infty)}$ is admissible for C'. Its improvement in the objective is exactly

$$2\varepsilon m(u) - \varepsilon^2(p_u + cp_u^2) > 0$$

for small enough ε. Hence $\alpha(C', \kappa) > \alpha(C, \kappa)$.

Continuity can be read directly from the unique root. If $C_j \to C > 1$, every subsequence of $u_{C_j,\kappa} \in [0, b]$ has a convergent further subsequence. Joint continuity of F in (C, u) makes its limit a zero of $F_{C,\kappa}$, necessarily $u_{C,\kappa}$. Thus the full sequence of roots converges; the formulas for ℓ and J give continuity of α.

For the lower endpoint, write an admissible profile as $q = s + h$ with $0 \le h \le (C-1)s$. Jensen's inequality gives $B_c(q, q) \ge (1+c)s^2$ for both signs of c. Since $\mathbb{E}Z = 0$ and $\mathbb{E}|Z| \le 1$,

$$\mathcal{V}_\kappa(q) \le 2(C-1)s - (1+c)s^2 \le \frac{(C-1)^2}{1+c}.$$

The zero profile is admissible, so $\alpha(C, \kappa) \to 0$ as $C \downarrow 1$.

For the upper endpoint, the residual proof for $p = (Z-\omega)_+$ in Section 10.1 gives $\mathcal{V}_\kappa(q) \le \mathcal{V}_\kappa(p) = \alpha_\star(\kappa)$ for every nonnegative $q \in L^2$. Conversely, let $B = b + |\omega|$ and, for $\varepsilon > 0$, set

$$q_\varepsilon(z) = \varepsilon + \min\{(z-\omega)_+, B\}.$$

If $C \ge 1 + B/\varepsilon$, this is admissible for C and equals $p + \varepsilon$ on the support of ν_{sc}. Direct expansion gives

$$\mathcal{V}_\kappa(q_\varepsilon) = \alpha_\star(\kappa) - (1+c)\big(2\varepsilon m(\omega) + \varepsilon^2\big).$$

Letting $\varepsilon \downarrow 0$ proves the limit $\alpha(C, \kappa) \to \alpha_\star(\kappa)$ as $C \to \infty$. Positivity for finite C follows from (70) and $q_\star \ge t > 0$. Strict monotonicity and the upper bound imply $\alpha(C, \kappa) < \alpha_\star(\kappa)$ at every finite C.

We have proved that $C \mapsto \alpha(C, \kappa)$ is a continuous strictly increasing bijection from $(1, \infty)$ onto $(0, \alpha_\star(\kappa))$. Its ordinary inverse is the stated infimum: if C_0 is the unique solution of $\alpha(C_0, \kappa) = \alpha$, the set in (28) is precisely $[C_0, \infty)$. This proves every inverse assertion of Theorem 3.3.

10.6 The specialization at fourth moment three

When $c = 0$, $L(u) = u/C$ and the root equation becomes $m(-t) = Cm(Ct)$ with $t = u/C \in (0, b/C)$. The derivative of $t \mapsto v(-t) + v(Ct)$ is $2\{m(-t) - Cm(Ct)\}$. Its sign changes exactly

once, from negative to positive, so $t = t_C$ is its unique minimizer on $[0, \infty)$. Consequently the supremum defining t_C agrees with the new root, $u_{C,3} = Ct_C$, and $\ell_{C,3} = t_C$.

At $c = 0$ the objective is $\mathbb{E}Z^2 - \mathbb{E}(Z - q)^2$. Minimizing the distance pointwise on $[t, Ct]$ gives

$$\alpha(C, 3) = \tfrac{1}{2} - v(-t_C) - v(Ct_C) = \alpha_C.$$

Equality of the defining sets gives equality of the inverses. To recover the stated global Lipschitz bound, let $1 < C \le C'$ and $s = t_{C'} < b/C'$. Testing the minimization for C at s yields

$$0 \le \alpha_{C'} - \alpha_C \le v(Cs) - v(C's) = 2\int_{Cs}^{C's} m(w)\, dw \le 2m(0)b(C' - C) \le 4(C' - C).$$

All remaining statements of Proposition 3.10 follow from the range and inverse already proved, with $\omega_3 = 0$ and $\alpha_\star(3) = v(0) = 1/4$.

10.7 The four endpoint asymptotics

Set $M = m(0) = 2\sqrt{2}/(3\pi)$ and $\rho = \sqrt{2}/\pi$, the maximum semicircle density. For $t \ge 0$, symmetry gives the useful identity

$$2(m(t) - M + t/2) = \mathbb{E}(t - |Z|)_+.$$

In particular,

$$0 \le m(t) - M + t/2 \le \rho t^2, \qquad |v(t + h) - v(t) + 2hm(t)| \le h^2. \tag{75}$$

The second bound follows by integrating $v' = -2m$ and the 1-Lipschitz bound for m. These estimates supply uniform remainders; no formal differentiation of an unspecified asymptotic series is needed.

As $C \downarrow 1$. Write $\eta = C - 1$ and $g_C(t) = v(t) - v(Ct) - t^2$. By (62), $\alpha_C = \max_{t \ge 0} g_C(t) = g_C(t_C)$. The root equation also gives the scale of its maximizer:

$$t_C = (C - 1)m(t_C) - C\{m(t_C) - m(Ct_C)\} \le M\eta.$$

Uniformly for $0 \le t \le M\eta$, (75) gives

$$g_C(t) = 2M\eta t - Ct^2 + O(\eta t^3 + \eta^2 t^2) = 2M\eta t - Ct^2 + O(\eta^4). \tag{76}$$

The quadratic on the right has its maximum at $t_0 = M\eta/C$, which also belongs to $[0, M\eta]$. Evaluating the lower bound at t_0 and the upper bound at t_C therefore proves

$$\alpha_{1+\eta} = \frac{M^2\eta^2}{1 + \eta} + O(\eta^4) = \frac{8}{9\pi^2}\eta^2(1 - \eta) + O(\eta^4).$$

The uniform constants in (75) make this exactly the existential remainder estimate of Theorem 3.11(i).

As $\alpha \downarrow 0$. Let $\eta = C_\star(\alpha) - 1$ and $\beta = \sqrt{\alpha}/M$. The inverse property makes $\eta \to 0$ and the preceding expansion gives

$$\beta^2 = \eta^2(1 - \eta) + O(\eta^4).$$

First $\beta/\eta \to 1$. Taking the positive square root, using the bounded second derivative of $\sqrt{1 + x}$ near zero, gives $\beta = \eta - \eta^2/2 + O(\eta^3)$. Thus $\eta - \beta = O(\beta^2)$, and substituting this once in the quadratic term gives $\eta = \beta + \beta^2/2 + O(\beta^3)$. Restoring M proves

$$C_\star(\alpha) = 1 + \frac{3\pi}{2\sqrt{2}}\sqrt{\alpha} + \frac{9\pi^2}{16}\alpha + O(\alpha^{3/2}).$$

This inversion uses only a neighborhood on which η and β are comparable, so the remainder has a finite constant as asserted.

As $C \to \infty$. Put $y_C = Ct_C \in (0, b)$. Then $t_C \le b/C \to 0$ and the root equation implies $Cm(y_C) = m(-t_C) \to M$. If y_C stayed below $b - \varepsilon$ along a subsequence, positivity of $m(b - \varepsilon)$ would contradict this boundedness. Hence $y_C \to b$. The exact decomposition of the deficit is

$$\tfrac{1}{4} - \alpha_C = v(-t_C) - v(0) + v(y_C).$$

The first term equals $2Mt_C + O(t_C^2)$ by (75). For the second, the pointwise bound $(Z - y)_+^2 \le (b - y)(Z - y)_+$ gives

$$0 \le Cv(y_C) \le (b - y_C)Cm(y_C) = (b - y_C)m(-t_C) \longrightarrow 0.$$

Since $Ct_C^2 \le b^2/C \to 0$, it follows that

$$C(\tfrac{1}{4} - \alpha_C) \longrightarrow 2Mb = \frac{8}{3\pi}.$$

Controlling the upper-tail term after multiplication by C is essential: its unscaled convergence to zero alone would not prove this limit.

As $\alpha \uparrow 1/4$. Strict monotonicity and $\alpha_C < 1/4$ at every finite C imply $C_\star(\alpha) \to \infty$. Substituting $C = C_\star(\alpha)$ and $\alpha_C = \alpha$ in the preceding limit gives

$$(\tfrac{1}{4} - \alpha)C_\star(\alpha) \longrightarrow \frac{8}{3\pi}.$$

This completes all four assertions of Theorem 3.11.

11 The Gaussian cone calculation

Fix $C > 1$ and $\kappa > 1$, and write $c = (\kappa - 3)/2$, $a_\kappa = \min\{1, 1 + c\} > 0$, and $b_\kappa = \max\{1, 1 + c\}$. This section proves the Gaussian input to the data theorem. The main step is a comparison of two quadratic maxima. Orthogonal invariance applies to one maximizer, although it does not apply to the original objective.

11.1 The covariance and the cone

Use the Frobenius scalar product on $\mathbb{S}_d$, and define

$$T_\kappa S = S_{\text{off}} + \sqrt{1 + c}\, S_{\text{diag}}, \qquad \mathcal{K}_C = \{S \in \mathbb{S}_d : tI_d \preceq S \preceq CtI_d \text{ for some } t \ge 0\}.$$

The cone $\mathcal{K}_C$ is closed and convex. Its trace slice $\mathcal{K}_C^{(d)} = \{S \in \mathcal{K}_C : \operatorname{tr} S = d\}$ is nonempty and compact, and every nonzero point of the cone has a unique positive multiple in this slice. Indeed, a nonzero matrix in $\mathcal{K}_C$ is positive definite. On the slice, $\lambda_{\max}(S) \le C$ and hence

$$d \le \|S\|_F^2 \le Cd, \qquad a_\kappa d \le \|T_\kappa S\|_F^2 \le b_\kappa Cd. \tag{77}$$

The second pair of inequalities follows from $\|T_\kappa S\|_F^2 = \|S\|_F^2 + c\|S_{\text{diag}}\|_F^2$.

Let $A_1, \dots, A_n$ be independent matrices with independent standard normal entries. The Gaussian features are specified by

$$\langle G_i, S\rangle = \sqrt{2/d}\, \langle A_i, T_\kappa S\rangle. \tag{78}$$

They have covariance $\mathbb{E}\langle G_i, S\rangle\langle G_i, R\rangle = d^{-1}(2\langle S, R\rangle + (\kappa - 3)\sum_j S_{jj}R_{jj})$, the covariance of the data features. The condition constraint is imposed on S, before T_κ is applied.

For one independent standard Gaussian matrix A, put

$$H = \frac{A + A^{\mathsf{T}}}{2\sqrt{d}}, \qquad \sigma_{C,\kappa}(A) = \sup_{S \in \mathcal{K}_C,\ \|T_\kappa S\|_F \le 1} \langle A, T_\kappa S \rangle.$$

Since T_κ is invertible, its image of $\mathcal{K}_C$ is a closed convex cone. The support function above is the norm of the projection onto that cone. The projection identity $\|\Pi_K z\|^2 = \max_{y \in K}(2\langle z, y\rangle - \|y\|^2)$, followed by the change of variable $S \mapsto \sqrt{d}\, S$, gives

$$\sigma_{C,\kappa}(A)^2 = d \sup_{S \in \mathcal{K}_C} \Phi_H(S), \qquad \Phi_H(S) = 2\langle H, T_\kappa S\rangle - \|S\|_F^2 - c\|S_{\mathrm{diag}}\|_F^2. \tag{79}$$

Proposition 11.1. *As $d \to \infty$,*

$$\frac{\sigma_{C,\kappa}(A)}{d} \xrightarrow{\mathbb{P}} \sqrt{\alpha(C,\kappa)}.$$

The proof occupies the next three subsections. The random-matrix inputs are the semicircle law and the spectral-edge bound in this normalization:

$$\frac{1}{d}\sum_{j=1}^{d} \delta_{\lambda_j(H)} \xrightarrow{\mathbb{P}} \nu_{\mathrm{sc}}, \qquad \mathbb{P}\{\|\sqrt{2}H\|_{\mathrm{op}} > 3\} \longrightarrow 0. \tag{80}$$

The first convergence means convergence of integrals of bounded continuous functions. We also use $\|H\|_F^2/d \to 1/2$ in probability, obtained directly by summing the independent squared Gaussian entries. These are the standard GOE semicircle and norm results; their formal proofs are in `library/random_matrix/goe.v` and `wigner.v`.

11.2 The equivariant maximizer

Replace the diagonal penalty by a trace penalty and set

$$\bar{\Phi}_H(S) = 2\langle H, S\rangle - \|S\|_F^2 - \frac{c}{d}(\operatorname{tr} S)^2, \qquad \mathcal{A}_c S = S + \frac{c}{d}(\operatorname{tr} S) I_d.$$

The operator $\mathcal{A}_c$ is positive definite, with smallest eigenvalue a_κ. Completing the square shows that $\bar{\Phi}_H$ has a unique maximizer $X(H)$ on $\mathcal{K}_C$:

$$\begin{aligned} \bar{\Phi}_H(S) &= \langle H, \mathcal{A}_c^{-1} H\rangle - \langle S - \mathcal{A}_c^{-1}H, \mathcal{A}_c(S - \mathcal{A}_c^{-1}H)\rangle, \\ \mathcal{A}_c^{-1}H &= H - \frac{c}{1+c}\,\frac{\operatorname{tr} H}{d} I_d. \end{aligned}$$

Thus the center of this weighted projection is $\mathcal{A}_c^{-1}H$. If $U_c = \mathcal{A}_c^{1/2}$, its explicit projection form is $X(H) = U_c^{-1}\Pi_{U_c\mathcal{K}_C}(U_c^{-1}H)$. Unlike T_κ, U_c changes only the identity direction.

The variational inequality is

$$\langle H - \mathcal{A}_c X, S - X\rangle \le 0 \qquad (S \in \mathcal{K}_C). \tag{81}$$

Taking $S = 0$ and $S = 2X$ gives $\langle H, X\rangle = \langle X, \mathcal{A}_c X\rangle$, so

$$\|X\|_F \le \|H\|_F / a_\kappa, \qquad \bar{\Phi}_H(X) = \langle H, X\rangle.$$

Adding the variational inequalities for $X(H)$ and $X(H')$ gives $a_\kappa\|X(H) - X(H')\|_F \le \|H - H'\|_F$. In particular, X is measurable. Orthogonal conjugation preserves the objective and the cone, so uniqueness gives

$$X(QHQ^{\mathsf{T}}) = QX(H)Q^{\mathsf{T}} \qquad (Q \text{ orthogonal}). \tag{82}$$

No choice of eigenvectors is involved.

Lemma 11.2. *If a symmetric random matrix Y has an orthogonally invariant law and $\mathbb{E}\|Y\|_F^2 < \infty$, then*

$$\mathbb{E}\|Y_{\mathrm{diag}} - (\operatorname{tr} Y/d)I_d\|_F^2 = \frac{2}{d+2}\mathbb{E}\left(\|Y\|_F^2 - \frac{(\operatorname{tr} Y)^2}{d}\right).$$

Consequently, for $Y = X(H)$,

$$\mathbb{E}\|X_{\mathrm{diag}} - (\operatorname{tr} X/d)I_d\|_F^2 \le \frac{d+1}{a_\kappa^2(d+2)}. \tag{83}$$

Proof. For $d = 1$ both sides vanish. For $d \ge 2$, permutations make the quantities $p = \mathbb{E}Y_{11}^2$, $q = \mathbb{E}Y_{11}Y_{22}$, and $r = \mathbb{E}Y_{12}^2$ independent of the chosen distinct indices. Sign changes give $\mathbb{E}(Y_{11} + Y_{22})Y_{12} = 0$. A rotation by $\pi/4$ in the first two coordinates makes $(Y_{11} + Y_{22})/2 + Y_{12}$ have the law of Y_{11}. Squaring gives $p = (p+q)/2 + r$, hence $p - q = 2r$. Now

$$\mathbb{E}\|Y\|_F^2 = dp + d(d-1)r, \qquad \mathbb{E}(\operatorname{tr} Y)^2 = dp + d(d-1)q.$$

Substitution proves the identity. The law of H is invariant under orthogonal conjugation, and (82) gives the same property for $X(H)$. Its norm bound and $\mathbb{E}\|H\|_F^2 = (d+1)/2$ prove (83). □

11.3 Comparison of the two maxima

Put $e = \sqrt{1+c} - 1$, $D = X_{\mathrm{diag}} - (\operatorname{tr} X/d)I_d$, and $E = eH_{\mathrm{diag}} - cD$. For $h = S - X$, expansion gives

$$\Phi_H(S) - \Phi_H(X) = 2\langle H - \mathcal{A}_c X, h\rangle + 2\langle E, h\rangle - \|h\|_F^2 - c\|h_{\mathrm{diag}}\|_F^2.$$

The last two terms are at most $-a_\kappa\|h\|_F^2$ for either sign of c. Apply (81) and complete a scalar square. Comparing the two objectives at X then gives the two bounds

$$\begin{aligned} \Phi_H(X) \le \sup_{S\in\mathcal{K}_C} \Phi_H(S) &\le \Phi_H(X) + \|E\|_F^2/a_\kappa, \\ \Phi_H(X) - \bar{\Phi}_H(X) &= 2e\langle H_{\mathrm{diag}}, X_{\mathrm{diag}}\rangle - c\|D\|_F^2. \end{aligned} \tag{84}$$

Controlling the value at X alone would give only one side of the comparison; the first line supplies the other side.

By (83), $\mathbb{E}\|D\|_F^2 = O(1)$, with constants depending on κ. Also $\mathbb{E}\|H_{\mathrm{diag}}\|_F^2 = 1$, so $\mathbb{E}\|E\|_F^2 = O(1)$ and Cauchy–Schwarz gives

$$\mathbb{E}|\langle H_{\mathrm{diag}}, X_{\mathrm{diag}}\rangle| \le \left(\mathbb{E}\|H_{\mathrm{diag}}\|_F^2\,\mathbb{E}\|X\|_F^2\right)^{1/2} = O(\sqrt{d}).$$

Divide (84) by d and use Markov's inequality:

$$\frac{1}{d}\left(\sup_{\mathcal{K}_C}\Phi_H - \sup_{\mathcal{K}_C}\bar{\Phi}_H\right) \xrightarrow{\mathbb{P}} 0. \tag{85}$$

The argument covers $-1 < c < 0$ directly; it does not use cone duality.

11.4 The shifted spectral problem

Let $V_d(H) = \sup_{\mathcal{K}_C}\bar{\Phi}_H$ and

$$P(B) = \sup_{S\in\mathcal{K}_C}(2\langle B, S\rangle - \|S\|_F^2) = \|B\|_F^2 - \operatorname{dist}_F(B, \mathcal{K}_C)^2.$$

For symmetric B, the spectral theorem and the Hoffman–Wielandt inequality give

$$\operatorname{dist}_F(B, \mathcal{K}_C)^2 = \min_{t \geq 0} \sum_{j=1}^{d} \operatorname{dist}(\lambda_j(B), [t, Ct])^2. \tag{86}$$

Indeed, for a fixed t, Hoffman–Wielandt bounds the distance below by the sum of the squared distances of the eigenvalues to $[t, Ct]$. Equality holds by retaining the eigenvectors of B and clipping its eigenvalues to that interval. Minimizing over t proves the formula.

Define $Q_d(\mu) = P(H - c\mu I_d) + cd\mu^2$. The exact identity

$$\bar{\Phi}_H(S) = 2\langle H - c\mu I_d, S\rangle - \|S\|_F^2 + cd\mu^2 - \frac{c}{d}(\operatorname{tr} S - d\mu)^2$$

shows that $V_d(H) \geq Q_d(\mu)$ if $c \leq 0$, and $V_d(H) \leq Q_d(\mu)$ if $c \geq 0$. These inequalities hold for every $\mu \in \mathbb{R}$. There is equality at $\mu_X = \operatorname{tr} X(H)/d$: by (81), $X(H)$ is the ordinary Frobenius projection of $H - c\mu_X I_d$ onto $\mathcal{K}_C$. Thus V_d is a maximum of Q_d when $c < 0$ and a minimum when $c > 0$.

Expanding the shifted norm before taking a limit gives

$$\frac{Q_d(\mu)}{d} = \frac{\|H\|_F^2}{d} - 2c\mu\frac{\operatorname{tr} H}{d} + c(1+c)\mu^2 - \inf_{t \geq 0} \frac{1}{d} \sum_{j=1}^{d} \operatorname{dist}(\lambda_j(H) - c\mu, [t, Ct])^2. \tag{87}$$

In particular the trace term has not been discarded by taking an expectation. We need convergence uniform at the random value μ_X.

Set $M = 1/a_\kappa$ and $B = 3 + |c|M$. On $\{\|H\|_F^2/d \leq 1\}$, positivity of X and its norm bound give $0 \leq \mu_X \leq M$. On $\{\|\sqrt{2}H\|_{\mathrm{op}} \leq 3\}$, all the numbers $\lambda_j(H) - c\mu$, for $0 \leq \mu \leq M$, are at most B. For $t > B$, all of them are below the lower endpoint t, so decreasing t to B cannot increase the clipping energy. The infimum in (87) may therefore be restricted to $[0, B]$. The same restriction applies to the limiting integral.

Here is the required uniform version of the semicircle limit. On the compact set $|z| \leq 3$, $0 \leq t \leq B$, $0 \leq \mu \leq M$, the functions

$$f_{\mu,t}(z) = \operatorname{dist}(z - c\mu, [t, Ct])^2 = (t + c\mu - z)_+^2 + (z - Ct - c\mu)_+^2$$

are uniformly bounded and have a common Lipschitz constant in (μ, t). This follows from $|x_+ - y_+| \leq |x - y|$ and $|x^2 - y^2| \leq |x - y|(|x| + |y|)$. Given $\eta > 0$, choose a finite (μ, t)-net whose mesh makes each approximation error at most $\eta/3$. At its finitely many points, the semicircle law gives convergence of the empirical integral in probability. Intersect those events and the spectral-norm event in (80). The triangle inequality then gives

$$\sup_{0 \leq \mu \leq M,\ 0 \leq t \leq B} \left| \frac{1}{d} \sum_j f_{\mu,t}(\lambda_j(H)) - \int f_{\mu,t}\, d\nu_{\mathrm{sc}} \right| \xrightarrow{\mathbb{P}} 0.$$

One can extend each function continuously outside $[-3, 3]$ by constants to apply the bounded-continuous form of the semicircle law. Taking an infimum in t cannot increase this uniform error. By symmetry,

$$\int f_{\mu,t}\, d\nu_{\mathrm{sc}} = v(-t - c\mu) + v(Ct + c\mu).$$

Finally $\mathbb{E}(\operatorname{tr} H/d)^2 = d^{-2}$, so the trace term in (87) vanishes uniformly on $[0, M]$. Together these estimates prove

$$\sup_{0 \leq \mu \leq M} \left| \frac{Q_d(\mu)}{d} - \mathcal{D}_{C,\kappa}(\mu) \right| \xrightarrow{\mathbb{P}} 0, \quad \mathcal{D}_{C,\kappa}(\mu) = \frac{1}{2} + c(1+c)\mu^2 - \inf_{t \geq 0}\{v(-t - c\mu) + v(Ct + c\mu)\}. \tag{88}$$

Let μ_0 be the mean of the scalar optimizer. The signed duality in (73) gives

$$0 \le \mu_0 \le \frac{1}{\sqrt{2}a_\kappa} < M, \qquad \mathcal{D}_{C,\kappa}(\mu_0) = \alpha(C,\kappa),$$

and $\mathcal{D}_{C,\kappa}(\mu) \le \alpha(C,\kappa)$ for all μ if $c \le 0$, with the reverse inequality if $c \ge 0$. For $c \le 0$, use $V_d \ge Q_d(\mu_0)$ for the lower bound and $V_d = Q_d(\mu_X)$ for the upper bound. For $c \ge 0$, use $V_d \le Q_d(\mu_0)$ for the upper bound and $V_d = Q_d(\mu_X)$ for the lower bound. Equation (88) therefore gives $V_d/d \to \alpha(C,\kappa)$ in probability, including $c = 0$. Equations (85) and (79) prove Proposition 11.1.

Formal source: `extension/ConditionKappa/GaussianLimit.v`, `kcond_support_cvg`.

The deterministic comparison is in `Geometry.v`; the weighted projection and rotation identities are in `library/matrix/trace_projection.v` and `library/random_matrix/goe_rotation.v`.

11.5 Gaussian comparison and its two consequences

We state the Gaussian comparison in the form used here. Let A be an $N \times p$ matrix with independent standard normal entries, and let $g \in \mathbb{R}^N$, $h \in \mathbb{R}^p$ be independent standard Gaussian vectors. For a nonempty bounded set $B \subset \mathbb{R}^p$ and $\rho > 0$ such that $\|x\| \le \rho$ on B, put

$$\mathcal{P}(\mathsf{A}) = \inf_{x \in B} \|\mathsf{A}x\|, \qquad \mathcal{Q}(g,h) = \inf_{x \in B} (\|g\|\|x\| - \langle h, x\rangle)_+.$$

For $N \ge 1$, the Gaussian min–max comparison gives

$$\mathbb{E}\mathcal{P} \ge \mathbb{E}\mathcal{Q} - \rho, \qquad \mathbb{E}\mathcal{P} \le \mathbb{E}\mathcal{Q} + \rho \quad \text{if } B \text{ is closed and convex.} \tag{89}$$

Moreover, for $u \ge 0$,

$$\begin{aligned} \mathbb{P}\{|\mathcal{P} - \mathbb{E}\mathcal{P}| \ge u\} &\le 2e^{-u^2/(2\rho^2)}, \\ \mathbb{P}\{|\mathcal{Q} - \mathbb{E}\mathcal{Q}| \ge u\} &\le 2e^{-u^2/(4\rho^2)}. \end{aligned} \tag{90}$$

These are standard Gaussian inputs, stated here to fix the normalization and the finite-dimensional error. For completeness, the comparison uses Gordon's inequality for

$$\langle y, \mathsf{A}x\rangle + \gamma\|x\|\|y\| \quad \text{and} \quad \|x\|\langle g, y\rangle - \|y\|\langle h, x\rangle, \qquad \|y\| \le 1,$$

where γ is an independent standard normal. Its extra term costs at most $\rho\mathbb{E}|\gamma| \le \rho$. Maximizing the auxiliary expression over the unit ball gives $(\|g\|\|x\| - \langle h, x\rangle)_+$. The upper bound uses convex minimax to exchange x and y before applying the comparison. Finite nets extend the finite-index Gordon inequality to the bounded sets. Finally $\mathcal{P}$ is ρ-Lipschitz in A and $\mathcal{Q}$ is $\sqrt{2}\rho$-Lipschitz in (g,h); Gaussian concentration proves (90). This is the mean-and-concentration form of the convex Gaussian min–max theorem [7, 18], proved in `library/comparison/cgmt.v`.

Apply this with the cone $K = T_\kappa\mathcal{K}_C$ and its base $B = T_\kappa\mathcal{K}_C^{(d)}$, expressed in full matrix Frobenius coordinates, so $p = d^2$. Its radii are

$$r = \sqrt{a_\kappa d}, \qquad \rho = \sqrt{b_\kappa C d}.$$

Write $\sigma_K(h) = \sup_{x \in K,\ \|x\| \le 1} \langle h, x\rangle$. Two deterministic facts connect this support function to $\mathcal{Q}$:

$$\begin{aligned} \|g\| < \sigma_K(h) \quad &\Longrightarrow \quad \mathcal{Q}(g,h) = 0, \\ &\mathcal{Q}(g,h) \ge r(\|g\| - \sigma_K(h))_+. \end{aligned} \tag{91}$$

For the first, the support maximum is attained on K intersected with the unit ball. Its maximizing point is nonzero under the strict inequality, and scaling it into B keeps $\|g\|\|x\| - \langle h, x\rangle$ negative. For the second, use $\langle h, x\rangle \le \|x\|\sigma_K(h)$ and $\|x\| \ge r$ on B.

Proposition 11.3 (The Gaussian witness)**.** *Suppose $n/d^2 \to \alpha$ and $0 < \alpha < \alpha(C, \kappa)$. For all sufficiently large d, with probability at least $3/4$, some $S \in \mathcal{K}_C^{(d)}$ satisfies*

$$\sum_{i=1}^{n} \langle G_i, S \rangle^2 \le 162 b_\kappa C. \tag{92}$$

Proof. Take $N = n$. The Gaussian norm law and Proposition 11.1 give $\|g\|/d \to \sqrt{\alpha}$ and $\sigma_K(h)/d \to \sqrt{\alpha(C,\kappa)}$ in probability. Thus $\mathbb{P}\{\mathcal{Q} = 0\} > 1/2$ eventually. The second concentration bound in (90), applied on this event, gives $\mathbb{E}\mathcal{Q} \le 2\rho\sqrt{\log 4}$. By (89), $\mathbb{E}\mathcal{P} \le \rho(1 + 2\sqrt{\log 4}) \le 5\rho$. The first concentration bound, with $u = 4\rho$, yields $\mathbb{P}\{\mathcal{P} \ge 9\rho\} \le 2e^{-8} < 1/4$. The compactness of B gives a minimizer. Transform it back to S and use (78): $\sum_i \langle G_i, S \rangle^2 < (2/d)(9\rho)^2 = 162 b_\kappa C$. □

Proposition 11.4 (The Gaussian gap)**.** *Suppose $n/d^2 \to \alpha > \alpha(C, \kappa)$, and set $\Delta = \sqrt{\alpha} - \sqrt{\alpha(C,\kappa)} > 0$. With probability tending to one, every $S \in \mathcal{K}_C^{(d)}$ and every $b \in \mathbb{R}$ satisfy*

$$\sum_{i=1}^{n} (\langle G_i, S \rangle - b)^2 \ge gn, \qquad g = \frac{a_\kappa \Delta^2}{64\alpha} > 0. \tag{93}$$

Proof. Let Q be an $n \times (n-1)$ matrix whose columns are an orthonormal basis of $\mathbf{1}^\perp$. For any $y \in \mathbb{R}^n$, $\inf_b \|y - b\mathbf{1}\|^2 = \|Q^\mathsf{T} y\|^2$. If A consists of the vectorized A_i, then $Q^\mathsf{T}\mathsf{A}$ has independent standard normal entries. Use the preceding comparison with $N = n - 1$. Centering removes one row; it does not change the limit of N/d^2.

We have $(\|g\| - \sigma_K(h))/d \to \Delta$ in probability. The nonnegative variable $(\|g\| - \sigma_K(h))_+$ is integrable, and eventually exceeds $\Delta d/2$ with probability at least $1/2$. Consequently (91) and (89) give $\mathbb{E}\mathcal{P} \ge r\Delta d/4 - \rho$. For large d, $\rho \le r\Delta d/16$. Taking $u = r\Delta d/16$ in the concentration bound proves

$$\mathbb{P}\{\mathcal{P} \le r\Delta d/8\} \le 2 \exp\left(-\frac{a_\kappa \Delta^2 d^2}{512 b_\kappa C}\right).$$

Outside this event, the squared centered Gaussian residual is at least

$$\frac{2}{d}\left(\frac{r\Delta d}{8}\right)^2 = \frac{a_\kappa \Delta^2 d^2}{32} \ge \frac{a_\kappa \Delta^2}{64\alpha}\, n,$$

where the last inequality holds eventually because $n \le 2\alpha d^2$. This controls all intercepts and all matrices simultaneously. □

The optimization values used above are Lipschitz functions of the Gaussian entries, so their exceptional events are Borel. The witness requires only a fixed probability of success and a bounded unnormalized residual; the gap has probability tending to one and is uniform in the intercept. These are the two different inputs needed for the data transfer.

Formal source: `extension/ConditionKappa/GaussianWitness.v`, `gaussian_kappa_witness`.

Formal source: `extension/ConditionKappa/GaussianGap.v`, `gaussian_kappa_gap`.

12 From Gaussian features to exact fits

Throughout this section, $\kappa > 1$, $K > 0$, and $n/d^2 \to \alpha > 0$ are fixed. We first treat finite mixtures. Countable mixtures will be obtained by a truncation that preserves the fourth moment exactly. The comparison and interpolation inputs from Koehler and Sohn [8] are stated below; the other established inputs are identified where used. We give explicitly the arguments extending these tools to mixtures and to a prescribed condition number.

12.1 Regularity survives mixing

Use the Frobenius inner product on $\mathbb{S}_d$ and put $W = (xx^{\mathsf{T}} - I_d)/\sqrt{d}$. Its covariance operator Σ_d is characterized, for symmetric A, B, by

$$\mathbb{E}\langle W, A\rangle\langle W, B\rangle = \frac{1}{d}\left(2\langle A, B\rangle + (\kappa - 3)\sum_{j=1}^{d} A_{jj}B_{jj}\right). \tag{94}$$

In particular, writing $m_\kappa = \min\{2, \kappa - 1\}$ and $M_\kappa = \max\{2, \kappa - 1\}$,

$$\frac{m_\kappa}{d}\|A\|_F^2 \le \langle A, \Sigma_d A\rangle \le \frac{M_\kappa}{d}\|A\|_F^2. \tag{95}$$

The feature dimension may equivalently be taken as d^2, with zero covariance on antisymmetric matrices. This is the convention of the formal proof.

For clarity, a centered feature Z with covariance Σ and $\lambda = \|\Sigma\|_{\mathrm{op}} > 0$ is called (K_f, τ)-regular if

$$\mathbb{E}\langle Z, A\rangle^4 \le K_f^4\lambda^2\|A\|_F^4, \qquad \mathrm{Var}\langle Z, TZ\rangle \le \tau^2\|T\|_F^2 \tag{96}$$

for every A and every self-adjoint operator T on the feature space. The second bound concerns quadratic functions of W, hence polynomials of degree four in the original point. It is not a tensorization assumption on the mixed point law.

Lemma 12.1. *For an admissible finite mixture with averaged fourth moment κ, W is centered, has covariance* (94)*, and is (K_f, τ_d)-regular with*

$$K_f = \frac{2^{10}K^2}{\sqrt{\kappa - 1}}, \qquad \tau_d^2 = \frac{(2^{31} + 16)K^8}{d}.$$

Moreover,

$$\mathbb{E}\exp\left(\frac{\sqrt{d}\,|\langle W, A\rangle|}{252K^2\|A\|_F}\right) \le 2 \quad (A \neq 0).$$

The same bounds are uniform in the number and weights of the components.

Proof. Condition on the label a and write $\kappa_{a,j} = \int z^4\,\mu_{d,a,j}(dz)$. Independence, centering, and unit variance within this component give

$$\mathbb{E}_a\langle W, A\rangle = 0, \qquad \mathbb{E}_a\langle W, A\rangle\langle W, B\rangle = \frac{1}{d}\left(2\langle A, B\rangle + \sum_j(\kappa_{a,j} - 3)A_{jj}B_{jj}\right).$$

Only paired indices survive the expansion; a coordinate appearing once has expectation zero. Averaging proves (94).

The independent-coordinate estimates underlying [8, Proposition 2.6] give, also when the coordinate fourth moments differ,

$$\mathbb{E}_a\langle W, A\rangle^4 \le \frac{9 \cdot 2^{33}K^8}{d^2}\|A\|_F^4, \qquad \mathrm{Var}_a\langle W, TW\rangle \le \frac{2^{31}K^8}{d}\|T\|_F^2.$$

These estimates use only the common subgaussian bound, not equality of the fourth moments. One way to obtain the second estimate is to resample a coordinate: the change of W is supported on its row and column, and the Efron–Stein inequality reduces the variance to the fourth moments of these changes. Summing their row and column weights counts each matrix entry at most twice. The first estimate follows by the same resampling argument for the quadratic form.

Let $\Sigma_{d,a}$ be the component covariance. The only nonzero entries of $\Sigma_{d,a} - \Sigma_d$ act on diagonal coordinates, and

$$\|\Sigma_{d,a} - \Sigma_d\|_F^2 = \frac{1}{d^2}\sum_j (\kappa_{a,j} - \kappa)^2 \le \frac{16K^8}{d}.$$

Here $0 \le \kappa_{a,j}, \kappa \le 4K^4$, by the exponential moment assumption. Total variance therefore gives

$$\begin{aligned}\operatorname{Var}\langle W, TW\rangle &= \sum_a w_a \operatorname{Var}_a\langle W, TW\rangle + \sum_a w_a \langle \Sigma_{d,a} - \Sigma_d, T\rangle^2 \\ &\le \frac{(2^{31}+16)K^8}{d}\|T\|_F^2.\end{aligned}$$

The fourth-moment bound averages directly. Since $\lambda_d \ge (\kappa - 1)/d$, its stated regularity constant follows. Finally, the component Hanson–Wright exponential bound averages directly too. None of these operations introduces the number of components into a constant. □

The Gaussian feature Z with covariance Σ_d is $(3^{1/4}, \sqrt{2}\lambda_d)$-regular: Gaussian fourth moments and $\operatorname{Var}\langle Z, TZ\rangle = 2\operatorname{tr}((\Sigma_d^{1/2} T \Sigma_d^{1/2})^2)$ give these bounds. Thus both feature laws satisfy (96) with the same bounded K_f and $\tau_d = O(d^{-1/2})$ after increasing the constants.

12.2 The comparison and interpolation inputs

Here is the quantitative comparison used in the proof. Let D be a nonempty, relatively open convex set with direction space V, and let $B \ge 0$ be a self-concordant function on D that tends to $+\infty$ at the relative boundary of D. For $\beta, \gamma > 0$, define

$$\Phi_Z = \inf_{S \in D,\ b \in \mathbb{R}} \left\{ \frac{1}{2n}\sum_{i=1}^n (\langle Z_i, S\rangle - b)^2 + \beta(\lambda_d \|S\|_F^2 + b^2) + \gamma B(S) \right\}. \tag{97}$$

For two (K_f, τ)-regular feature laws with the same covariance, put

$$L = (1 + \beta^{-1})\left(1 + \inf_{S \in D}\{\lambda_d \|S\|_F^2 + \gamma B(S)\}\right).$$

Theorem 2.9 of Koehler and Sohn [8], in its form proved in the library, gives, when $n\gamma \ge 36$,

$$|\mathbb{E}\Phi_W - \mathbb{E}\Phi_Z| \le 100(1 + K_f^4)^2 L^2 \left\{ (n\gamma)^{-1/2} + \frac{\sqrt{\dim V}}{n}\left(\frac{\tau}{\lambda_d} + 1\right)\right\} =: \mathcal{E}_d, \tag{98}$$

$$\operatorname{Var}\Phi_W, \operatorname{Var}\Phi_Z \le \frac{4(1 + K_f^4)^2 L^2}{n}. \tag{99}$$

In particular, both values concentrate around the same deterministic center whenever $\mathcal{E}_d \to 0$. The proof replaces feature rows one at a time, bounds the insertion cost at the minimizer with that row removed, and uses the quadratic-form variance bound in (96) for its random Hessian term. This is why the covariance-spread estimate above is sufficient for mixtures.

We also use the independent-array form of the nuclear-norm estimate behind [8, Theorems 2.3 and 1.9]. The paper states its interpolation conclusion for identically distributed rows and residuals in the range; the formal proof establishes the following stronger input and conclusion. If all nd entries of the data array are independent, centered, of variance one, and have the same subgaussian bound K, then, for $0 < \alpha < 1/2$, there is $L_I = L_I(\alpha, K)$ such that, outside a Borel exceptional set of arbitrarily small probability, every $y \in \mathbb{R}^n$ admits symmetric Δ and b with

$$\langle W_i, \Delta\rangle = y_i + b, \qquad \|\Delta\|_{\mathrm{op}} \le L_I\left(\frac{\|y\|_2}{d} + \frac{(\log d)^4}{\sqrt{d}}\|y\|_\infty\right). \tag{100}$$

The eventual dimension bound is uniform over the coordinate laws, which need not agree across rows. This uniformity is needed below. Here is its precise analytic input. With high probability, uniformly over such independent arrays, every $\lambda \in \mathbb{R}^n$ has a decomposition $\lambda = h + t$ on disjoint coordinate sets such that

$$\left\| \sum_i \lambda_i W_i \right\|_* \geq c_I \left(\frac{\sqrt{d}}{(\log d)^4} \|h\|_1 + d\|t\|_2 \right), \tag{101}$$

where $c_I > 0$ depends only on α, K and $\|\cdot\|_*$ is the nuclear norm. This is the form of the published nuclear-norm argument used here. Its estimates for sparse vectors and for tails with small individual coordinates are uniform in the coordinate laws: they use only the common subgaussian bound and the scalar anti-concentration consequence of mean zero and variance one. The threshold $\alpha < 1/2$ enters the latter estimate through the number of independent off-diagonal constraints. These estimates are proved for independent arrays in `extension/Sat/SparseLower.v` and `extension/Sat/LightTailBound.v`; shelling the largest coordinates gives (101).

For completeness, its deterministic implication is short and removes the range restriction. For $\sum_i \lambda_i = 0$,

$$|\langle \lambda, y \rangle| \leq \|h\|_1 \|y\|_\infty + \|t\|_2 \|y\|_2 \leq \frac{1}{c_I} \left(\frac{(\log d)^4}{\sqrt{d}} \|y\|_\infty + \frac{\|y\|_2}{d} \right) \left\| \sum_i \lambda_i W_i \right\|_* .$$

Put $P = I_n - \mathbf{1}\mathbf{1}^\mathsf{T}/n$, and let $\mathcal{A}\Delta$ be the orthogonal projection of $(\langle W_i, \Delta \rangle)_i$ onto $\mathbf{1}^\perp$. If Py were outside the compact convex image under $\mathcal{A}$ of the operator-norm ball of the displayed radius, separation would give a $\lambda \in \mathbf{1}^\perp$ violating this inequality: the support function of that ball is the nuclear norm of $\mathcal{A}^*\lambda = \sum_i \lambda_i W_i$. Hence $\mathcal{A}\Delta = Py$ for some matrix in the ball, which is exactly (100).

To pass this input to finite mixtures, fix a dimension and condition on all row labels $s = (a_1, \ldots, a_n)$. For each s, take a Borel set E_s of conditional probability at most ε containing every sample for which (100) fails. The finite intersection $E = \bigcap_s E_s$ still contains every failure. For each s, $P_s(E) \leq P_s(E_s) \leq \varepsilon$, and hence

$$Q_d^{\otimes n}(E) = \sum_s \left(\prod_i w_{a_i} \right) P_s(E) \leq \varepsilon.$$

Thus neither a union bound over the labels nor measurability of the intrinsic failure set is required. This proves Theorem 2.7 for finite mixtures, with the same L_I.

12.3 Finite spectral covers and approximate fits

Write

$$\mathcal{S}_C = \mathcal{K}_C^{(d)} = \{S \in \mathcal{K}_C : \operatorname{tr} S = d\}, \qquad D_{a,b} = \{S \in \mathbb{S}_d : \operatorname{tr} S = d,\ aI_d \prec S \prec bI_d\}.$$

For $0 < a < 1 < b$, the normalized logarithmic barrier on $D_{a,b}$ is

$$B_{a,b}(S) = -\log\det(S - aI_d) - \log\det(bI_d - S) + d\log((1-a)(b-1)). \tag{102}$$

It is nonnegative, by concavity of the logarithm and $\operatorname{tr} S = d$, vanishes at I_d, and is self-concordant.

Lemma 12.2 (A cover with fixed spectral margins)**.** *For $1 \leq C_0 < C_1$, there are finitely many pairs (a_j, b_j) and $B_0 < \infty$, independent of d, such that*

$$0 < a_j < 1 < b_j < C_1 a_j,$$

and every $S \in \mathcal{S}_{C_0}$ belongs to some D_{a_j,b_j} with $B_{a_j,b_j}(S) \leq B_0 d$.

Proof. Such an S has its spectrum in $[t, C_0 t]$ for some $t \in [C_0^{-1}, 1]$. Put

$$h = \frac{(C_1 - C_0)C_0^{-1}}{2(C_1 + C_0 + 1)}, \qquad u_j = C_0^{-1} + jh \quad \left(0 \le j \le \left\lfloor \frac{1 - C_0^{-1}}{h} \right\rfloor\right).$$

Choose j with $u_j \le t < u_j + h$, and take $a_j = u_j - h$, $b_j = C_0(u_j + h) + h$. The displayed inequalities follow from the choice of h. Every eigenvalue is at distance at least h from both endpoints of this slice. Consequently

$$B_{a_j, b_j}(S) \le d\left[\log\{(1+h)(C_0(1+h)+h)\} - 2\log h\right].$$

The number of slices and this bound do not depend on d. □

On any of these slices use (97) with $\gamma_d = d^{-3/2}$ and a fixed $\beta > 0$. Then $L \le (1+\beta^{-1})(1 + M_\kappa)$ by evaluation at I_d. Since $\dim V \le d^2$, $\lambda_d \ge (\kappa - 1)/d$, and $\tau_d = O(d^{-1/2})$, the error in (98) is $O(d^{-1/4})$. Both values therefore differ by $o_{\mathbb{P}}(1)$ from $c_{j,d} = \mathbb{E}\Phi_{Z,j}$, uniformly over a fixed finite cover. Here and below β is fixed before letting d tend to infinity.

Suppose Proposition 11.3 at C_0 gives, with probability at least $3/4$, some $S \in \mathcal{S}_{C_0}$ satisfying $\sum_i \langle Z_i, S\rangle^2 \le B_G$, where B_G is fixed. By Lemma 12.2, on that event some slice j has

$$\Phi_{Z,j} \le \frac{B_G}{2n} + \beta_j M_\kappa b_j + B_0 d^{-1/2}.$$

The slice may depend on the Gaussian sample. To obtain a deterministic choice, choose any fixed positive tolerances for the finitely many values $\Phi_{Z,j}$. Their simultaneous concentration event has probability tending to one, so it intersects the witness event. At a point in this intersection, the slice containing the witness satisfies the same upper bound for its deterministic center, up to its tolerance. Thus, for every sufficiently large d, at least one index $j = j(d)$ satisfies that bound. Applying comparison on this deterministic slice is now legitimate.

We record the residual control needed for interpolation. For every $\delta > 0$, choose

$$A_j = M_\kappa b_j + 1, \qquad \beta_j = \frac{\delta}{2A_j}.$$

Take the concentration tolerances and then $d^{-1/2}B_0$ small compared with β_j. The preceding choice of slice and comparison give, with probability tending to one, $\Phi_{W,j} \le \beta_j A_j$. Its minimizer (S, b) consequently has

$$\sum_i (\langle W_i, S\rangle - b)^2 \le \delta n. \tag{103}$$

The minimizer exists: the barrier diverges at the spectral boundary, and the ridge controls the intercept. The positive ridge makes this minimizer and each minimizer with one row removed unique Borel functions of their input rows.

The bound on the largest residual requires a separate argument. Let $(S^{(i)}, b^{(i)})$ minimize the same objective after removing the ith loss summand, retaining its factor $1/(2n)$ and both regularizers. Comparison of the two minimizing inequalities gives

$$|\langle W_i, S\rangle - b| \le |\langle W_i, S^{(i)}\rangle - b^{(i)}|.$$

On the event $\Phi_{W,j} \le \beta_j A_j$, the objective with one row removed is no larger, hence $|b^{(i)}| \le \sqrt{A_j}$. Also $\|S^{(i)}\|_F \le \sqrt{b_j d}$ throughout the slice. The removed row is independent of $S^{(i)}, b^{(i)}$; the exponential bound of Lemma 12.1 yields

$$\mathbb{P}\left(|\langle W_i, S^{(i)}\rangle| > 3 \cdot 252 K^2 \sqrt{b_j}\log d\right) \le 2d^{-3}.$$

Taking the union over $n = O(d^2)$ rows proves

$$\max_i |\langle W_i, S\rangle - b| \le A \log d \tag{104}$$

with probability tending to one. The constant A can be maximized over the finite cover. This proves the approximate-fit assertion with both norms of the residual controlled; an average loss bound alone would not suffice.

12.4 The two sides of the prescribed-condition threshold

Assume first $\alpha < \alpha(C, \kappa)$. By continuity of the curve, choose $1 < C_- < C$ with $\alpha < \alpha(C_-, \kappa)$. Proposition 11.3 supplies B_G for C_-. Apply the cover with $(C_0, C_1) = (C_-, C)$ and set

$$e = \min_j \frac{Ca_j - b_j}{C + 1} > 0.$$

Then $e < a_j$ and $(b_j + e)/(a_j - e) \le C$ for every j. Since $\alpha < \alpha_\star(\kappa) < 1/2$, interpolation is available. For $y_i = b - \langle W_i, S\rangle$, equations (103)–(104) and (100) give

$$\|\Delta\|_{\mathrm{op}} \le L_I \left(\sqrt{\delta}\, \frac{\sqrt{n}}{d} + A \frac{(\log d)^5}{\sqrt{d}} \right).$$

Choose δ so that $L_I \sqrt{2\alpha\delta} < e/2$ and then take d large. The correction has operator norm at most e and

$$R = S + \Delta \succ 0, \qquad \operatorname{cond} R \le \frac{b_j + e}{a_j - e} \le C, \qquad \langle W_i, R\rangle = b + b'$$

for a common b'. Hence every $x_i^{\mathsf{T}} R x_i$ has the same value q. This value is positive as soon as one point is nonzero. For each component and coordinate, Cauchy–Schwarz gives $\mathbb{P}(x_j \neq 0) \ge 1/\mathbb{E}x_j^4 \ge 1/(4K^4)$. Consequently the probability that every point is zero is at most $(1 - 1/(4K^4))^n \to 0$. Dividing R by q yields an exact fit at level one without changing its condition number. This proves the subcritical assertion.

Now suppose $\alpha > \alpha(C, \kappa)$. Choose $C_+ > C$ with $\alpha > \alpha(C_+, \kappa)$ and cover $\mathcal{S}_C$ by slices contained in $\mathcal{S}_{C_+}$. Proposition 11.4 gives a fixed $g > 0$ such that, with probability tending to one,

$$\sum_i (\langle Z_i, S\rangle - b)^2 \ge gn \quad (S \in \mathcal{S}_{C_+},\ b \in \mathbb{R}). \tag{105}$$

Each Gaussian barrier value is then at least $g/2$. Its center is at least $g/4$ for all large d, because the value is nonnegative. Comparison therefore makes every data barrier value at least $g/8$ with probability tending to one, for any fixed positive ridge weights on the finitely many slices.

An exact fit of condition number at most C, normalized to trace d, would give $R \in \mathcal{S}_C$ and a common value $b_0 = \langle W_i, R\rangle$. On its covering slice,

$$\Phi_{W,j} \le \beta_j (M_\kappa b_j + b_0^2) + B_0 d^{-1/2}. \tag{106}$$

It remains to control b_0; the fit event alone puts no deterministic bound on it. Write $\overline{W} = n^{-1} \sum_i W_i$. For every such R,

$$b_0^2 = \langle \overline{W}, R\rangle^2 \le b_j d \|\overline{W}\|_F^2, \qquad \mathbb{E}[d\|\overline{W}\|_F^2] = \frac{d \operatorname{tr} \Sigma_d}{n} = O(1).$$

Thus for any $\varepsilon > 0$ there is $H_\varepsilon < \infty$ such that, outside an event of probability at most ε, $b_0^2 \le b_j H_\varepsilon$ simultaneously for all possible fits. Choose the finitely many $\beta_j > 0$ so that

$\beta_j b_j(M_\kappa + H_\varepsilon) < g/32$, then take d large in (106). Its upper bound is below $g/8$, a contradiction. Hence the fit probability has limsup at most ε, and $\varepsilon \downarrow 0$ proves the supercritical assertion. This argument imposes no upper bound on α.

The formal proof uses the stronger uniform estimate $\|\sum_i y_i W_i\|_F \le C_a\sqrt{d}\|y\|_2$ to bound the same intercept. The second-moment argument above is sufficient for the probability limit and makes the order of choices explicit.

12.5 Measurability and the optimal condition number

For fixed n, d, C, the conditioned-fit event is Borel. Indeed, restrict the matrix to $\|S\|_F \le m$ and $S \succeq m^{-1} I_d$. The equations $x_i^{\mathsf{T}} S x_i = 1$ and the condition-number inequality are closed conditions. Projection over this compact set of matrices is closed in the sample; taking the union over positive integers m gives $\mathcal{F}_{n,d}(C)$. The same argument without the condition-number inequality gives the positive definite fit event, and with $S \succeq 0$ gives the positive semidefinite fit event.

No attainment of the infimum defining $\mathcal{C}$ is needed. Monotonicity of the events gives, for every real r,

$$\{\mathcal{C} < r\} = \bigcup_{q \in \mathbb{Q},\ q<r} \mathcal{F}_{n,d}(q). \tag{107}$$

This proves extended-real measurability, including the value $+\infty$.

Let $C_* = C_\star(\alpha, \kappa)$ for $0 < \alpha < \alpha_\star(\kappa)$. For any $\varepsilon > 0$, choose

$$1 < C_- < C_* < C_+, \qquad C_* - \varepsilon < C_-, \quad C_+ < C_* + \varepsilon.$$

Strict monotonicity and the inverse identity give $\alpha(C_-, \kappa) < \alpha < \alpha(C_+, \kappa)$. With probability tending to one, there is no fit at C_- and there is a fit at C_+. On that event $C_- \le \mathcal{C} \le C_+$, which proves convergence in probability. The union of the two exceptional Borel events also gives exactly the exceptional-set formulation of the frozen statement.

12.6 The unrestricted fitting error

We give the remaining transfer argument for Theorem 2.5. Write

$$\Psi_W = \inf_{S \succeq 0,\ \operatorname{tr} S = d,\ b \in \mathbb{R}} \frac{1}{2n} \sum_i (\langle W_i, S\rangle - b)^2.$$

For data features, $2\Psi_W = \Gamma(X)$. In fact $r = \sqrt{d} + b$ converts the objectives, and the optimal fitted level r is the mean of the nonnegative numbers $x_i^{\mathsf{T}} S x_i/\sqrt{d}$, so its nonnegativity constraint costs nothing. Positive definite matrices are dense in this trace slice by $S \mapsto (1-\eta)S + \eta I_d$; the same infimum is therefore obtained over its interior.

The Gaussian error theorem [8, Proposition 1.4(ii)] states, for Gaussian features of covariance (94),

$$2\Psi_Z \longrightarrow e_\star(\alpha, \kappa) \quad \text{in probability}, \qquad \alpha > 0. \tag{108}$$

The cutoff and value in this formula are those of Definition 2.1. The argument centers the rows to eliminate b, applies the Gaussian min–max comparison on the positive semidefinite trace slice, and evaluates its scalar minimum by the semicircle law. This established Gaussian theorem is the input for the unrestricted error; the prescribed-condition result uses the different cone computation proved above.

To compare Ψ_W and Ψ_Z, a fixed ridge is insufficient: the trace slice has matrices with Frobenius norm of order d. Use the barrier $B(S) = -\log\det S$ on its positive definite interior and, for $d > e$, take

$$\beta_d = (\log d)^{-5/4}, \qquad \gamma_d = \frac{1}{d(\log d)^2}. \tag{109}$$

Here $B \geq 0$ and $B(I_d) = 0$. In (98), $L = O(\beta_d^{-1})$, $\lambda_d \asymp d^{-1}$, and $\tau_d = O(d^{-1/2})$, so the comparison error and the variance both vanish. For example, the first error term is $O(d^{-1/2}(\log d)^{7/2})$, and the covariance term is $O(d^{-1/2}(\log d)^{5/2})$.

Removing the barrier is controlled by contracting toward I_d. The self-concordant barrier bound gives, with contraction $\eta_d = d^{-1}$, an extra cost at most

$$\eta_d F_{Z,\beta_d}(I_d, 0) + \gamma_d d \log(1/\eta_d), \qquad F_{Z,\beta_d}(I_d, 0) = \frac{1}{2n} \sum_i \langle Z_i, I_d \rangle^2 + \beta_d \lambda_d \, d.$$

The same notation with W replaces Z for data features. The expectation of this reference objective is bounded, so the barrier cost tends to zero in probability for either feature law. Let Ψ_{W,β_d} denote the infimum defining Ψ_W with the added ridge $\beta_d(\lambda_d \|S\|_F^2 + b^2)$. The ridge-removal estimate of [8, Lemma 2.12] applies because (95) is coercive on symmetric matrices and the rank-one forms $\langle W, vv^{\mathsf{T}} \rangle$ have subexponential norm $O(d^{-1/2})$. It gives

$$0 \leq \Psi_{W,\beta_d} - \Psi_W \leq \left(1 + \frac{M_\kappa}{m_\kappa}\right) \beta_d \log d \tag{110}$$

outside an event of probability tending to zero; the same holds for Z. The right-hand side is $O((\log d)^{-1/4})$. More precisely, the localization behind this estimate bounds $\langle S, \Sigma_d S \rangle + b^2$ for all near-minimizers by a constant chosen for a prescribed probability tolerance; for large d that constant is at most $\log d$.

Consequently both unregularized values concentrate around the same deterministic sequence. By (108) that sequence tends to $e_\star(\alpha, \kappa)/2$, proving the error limit for every $\alpha > 0$, including $\alpha = \alpha_\star(\kappa)$. The statistic is measurable: minimize the continuous objective, with the common level already eliminated, over the compact positive semidefinite trace slice.

Above $\alpha_\star(\kappa)$ the error limit is positive. Any exact positive semidefinite fit, normalized to trace d, has zero error; thus its event is contained in $\{|\Gamma - e_\star| > e_\star/2\}$ and has probability tending to zero. Below $\alpha_\star(\kappa)$, choose a fixed $C > 1$ with $\alpha < \alpha(C, \kappa)$, using the limit of the curve as $C \to \infty$. The prescribed-condition theorem then supplies both the positive definite fit and the bounded-condition fit. The choice of C precedes the mixture family and the sample-size sequence. This completes all three assertions for finite mixtures.

12.7 Countable mixtures and total variation

Choose a cutoff M_d so that the discarded label mass $r_d = \sum_{a > M_d} w_{d,a}$ is at most $(d+1)^{-3}$. Keep the first $M_d + 1$ components. Replace the tail by one product component of weight r_d whose jth coordinate law is

$$\widetilde{\mu}_{d,j} = r_d^{-1} \sum_{a > M_d} w_{d,a} \mu_{d,a,j} \quad (r_d > 0).$$

If $r_d = 0$, use any existing admissible component law in that zero-weight slot. This finite mixture $\widetilde{Q}_d$ has the same K: mean, variance, and exponential moment bounds pass through the coordinate mixtures. More importantly,

$$r_d \int z^4 \, \widetilde{\mu}_{d,j}(dz) = \sum_{a > M_d} w_{d,a} \int z^4 \, \mu_{d,a,j}(dz),$$

so its averaged fourth moment is exactly κ. Simply deleting and renormalizing the tail would not preserve this constraint.

The two point laws agree on the retained mixture part; therefore $\mathrm{TV}(Q_d, \widetilde{Q}_d) \leq r_d$. Coupling the retained labels in each row, or telescoping the product measures, gives

$$\mathrm{TV}(Q_d^{\otimes n}, \widetilde{Q}_d^{\otimes n}) \leq n r_d = o(1).$$

All measurable fit and error probabilities transfer immediately. For interpolation and the optimum, take the exceptional Borel set provided for the finite family; its probability increases by at most nr_d and its deterministic conclusion is unchanged. The constants C, L_I are those of the finite theorem. No rate of decay of the original weights is needed, because M_d may vary freely with d.

For completeness, the exact identity (20) follows first on rectangles by expanding the product of the nonnegative component sums, and then on all measurable sets by uniqueness of product measures. Finite n and nonnegative summands justify exchanging the sums. For arbitrary probability laws P, Q, a telescoping replacement of one factor at a time gives

$$|P^{\otimes n}(A) - Q^{\otimes n}(A)| \le n \operatorname{TV}(P, Q)$$

for every measurable A. Indeed, integrating the section of A over the unchanged factors leaves a measurable function with values in $[0, 1]$, whose integrals differ by at most $\operatorname{TV}(P, Q)$. If $d^2 \operatorname{TV}(P_d, Q_d) \to 0$, then

$$n \operatorname{TV}(P_d, Q_d) = \frac{n}{d^2}\, d^2 \operatorname{TV}(P_d, Q_d) \longrightarrow 0.$$

This proves all robustness conclusions, including interpolation and the optimum, by using their same measurable exceptional sets. The argument involves no moments of P_d. The newer robust condition-number proof uses this exact bound; the older robustness assembly may also be proved through the preceding truncation.

For the contamination proposition, independently mark each row for replacement by zero with probability $p = \lambda/n$. If at least one row is marked, a fit at level one is impossible. Conditional on no mark, the sample has law $Q_d^{\otimes n}$. Thus

$$(P_d^{(\lambda)})^{\otimes n}(\mathcal{F}^{\mathrm{pd}}) = (1 - \lambda/n)^n Q_d^{\otimes n}(\mathcal{F}^{\mathrm{pd}}).$$

Also $\operatorname{TV}((1-p)Q_d + p\delta_0, Q_d) \le p$. The subcritical fit probability tends to one and $(1-\lambda/n)^n \to e^{-\lambda}$, proving both assertions. The identity remains valid when Q_d itself has an atom at zero.

12.8 The original Gaussian conclusions

The Gaussian point law is an admissible one-component family with $\kappa = 3$. The fixed-density transition therefore follows from Theorem 2.5. If $\limsup n(d)/d^2 < 1/4$, choose ρ strictly between this limsup and $1/4$ and couple the sample with its first $n(d)$ rows inside a sample of size $\lceil \rho d^2 \rceil$. A fit of the larger sample fits the smaller one. If $\liminf n(d)/d^2 > 1/4$, use instead a subsample of size $\lfloor \rho d^2 \rfloor$ with $1/4 < \rho < \liminf n(d)/d^2$. These monotonicities prove the one-sided-density statements, even when the density has no limit.

The prescribed-trace conclusion needs an additional Gaussian input: the common level produced by interpolation above need not equal d when the trace is d. We use the fixed-density form of [10, Theorems 1.3 and 3.1]: for every $0 < \rho < 1/4$ there are $0 < \lambda_- \le \lambda_+$ such that, for $n = \lfloor \rho d^2 \rfloor$, with probability tending to one there exists S satisfying simultaneously

$$\operatorname{tr} S = d, \qquad \lambda_- I_d \preceq S \preceq \lambda_+ I_d, \qquad x_i^{\mathsf{T}} S x_i = d \quad (1 \le i \le n).$$

This Gaussian input, including its proof by the feature spectral edges and the head–tail separation argument, is also formalized independently in the original part of the repository. Choose ρ with $\max\{\beta, 0\} < \rho < 1/4$ and delete rows as above. The constants then depend only on β, giving Theorem 7.2. If $\beta < 0$, there is no sample-size sequence satisfying its density hypothesis, so the probability assertion is vacuous. Compact projection proves measurability of the prescribed-trace event.

Finally, let $L = \{\sum_i y_i x_i x_i^{\mathsf{T}} : \sum_i y_i = 0\} \subset \mathbb{S}_d$. The semidefinite alternative says that exactly one of

$$L \cap \{H \succ 0\} \neq \varnothing, \qquad L^{\perp} \cap \{S \succeq 0,\ S \neq 0\} \neq \varnothing$$

holds. To see existence when the first set is empty, separate the linear space L from the open positive definite cone. The separating functional vanishes on L, is nonnegative on that cone and hence on its closure, and is represented by a nonzero positive semidefinite matrix. Mutual exclusion follows from $\operatorname{tr}(SH) > 0$ for $S \succeq 0$, $S \neq 0$, and $H \succ 0$.

If the points span $\mathbb{R}^d$, a matrix in the second set has $x_i^{\mathsf{T}} S x_i = q$ independent of i, and $q > 0$: otherwise $S^{1/2} x_i = 0$ for all i, forcing $S = 0$. It can therefore be rescaled to an exact fit. Gaussian points span almost surely when $n \geq d$, since the determinant of the first d points is a nonzero polynomial and has zero set of Lebesgue measure zero. Below the threshold a fit always excludes a balanced positive definite combination. Above it, $n \geq d$ eventually, and absence of a fit implies such a combination almost surely. The transition already proved gives Theorem 7.3. Its event is empty when $n = 0$. For $n \geq 1$ it is Borel: positive definiteness is open, so it suffices to take rational $y_1, \dots, y_{n-1}$ and set $y_n = -\sum_{i<n} y_i$.

13 Proofs for the dependent laws

13.1 Product-mixture representations

The polynomial identities needed for both examples can be recorded at once. With $X \sim \mu$ and h as in (47), integration of monomials on $[-\sqrt{3}, \sqrt{3}]$ gives

$$\begin{gathered} \mathbb{E}h(X) = \mathbb{E}[Xh(X)] = \mathbb{E}[X^2 h(X)] = 0, \qquad \mathbb{E}[X^4 h(X)] = \frac{8}{35}, \\ \mathbb{E}h(X)^2 = \frac{1}{9}, \qquad \mathbb{E}|X| = \frac{\sqrt{3}}{2}, \qquad \mathbb{E}[|X|h(X)] = -\frac{\sqrt{3}}{48}. \end{gathered} \tag{111}$$

For example, these follow by substituting $\mathbb{E}X^{2j} = 3^j/(2j+1)$; for the last identity use the evenness of $|x|h(x)$ and integrate $xh(x)$ on $[0, \sqrt{3}]$. The stationary points of $P_4(y) = (35y^4 - 30y^2 + 3)/8$ on $[-1, 1]$ are 0 and $\pm\sqrt{3/7}$. Evaluation at these points and the endpoints gives $-3/7 \leq h \leq 1$. Consequently $1 + ah \geq 0$ for $a \in [-1, 7/3]$, its integral is one, and (111) gives the moments of λ_a. Its support also gives $\int e^{x^2/9}\, d\lambda_a \leq e^{1/3} < 2$.

The important point for the paired law is a representation that remains admissible at both ends of the entire parameter interval. If $-1 \leq \epsilon < 0$, put $s = \sqrt{-\epsilon}$. Then

$$1 + \epsilon h(x)h(y) = \frac{1}{2}(1 + sh(x))(1 - sh(y)) + \frac{1}{2}(1 - sh(x))(1 + sh(y)). \tag{112}$$

If $0 \leq \epsilon \leq 7/3$, use instead

$$\begin{aligned} 1 + \epsilon h(x)h(y) = {} & \frac{\epsilon}{1+\epsilon}(1 - h(x))(1 - h(y)) \\ & + \frac{1}{1+\epsilon}(1 + \epsilon h(x))(1 + \epsilon h(y)). \end{aligned} \tag{113}$$

All parameters in these factors belong to $[-1, 7/3]$. In each coordinate their weighted mean is zero, so their averaged fourth moment is $9/5$. Multiply the two-component expansions independently over $\lfloor d/2 \rfloor$ pairs and, for odd d, append λ_0. This gives a finite product mixture with at most $2^{\lfloor d/2 \rfloor}$ components, the common constant $K = 3$, and averaged fourth moment $9/5$. Zero weights at $\epsilon = 0$ cause no problem because their component laws remain admissible. Theorem 2.5 and Theorem 2.7 now give Theorem 4.2, with constants independent of ϵ. For the latent-sign law, the defining two-component representation already has the same K and averaged fourth moment, proving Theorem 4.4.

The claimed range of the paired density is exact. The product $h(x)h(y)$ ranges from $-3/7$ to 1. Thus nonnegativity requires $1 + \epsilon \geq 0$ and $1 - 3\epsilon/7 \geq 0$, and these conditions suffice.

13.2 The latent covariance identities

Let B be the latent sign, with values ± 1 of equal probability. Given B, coordinates are independent with law λ_{Bs}. Since $(\lambda_s + \lambda_{-s})/2 = \mu$, every marginal is μ in every positive dimension. By (111),

$$\mathbb{E}[h(X_j) \mid B] = \frac{Bs}{9}, \qquad \mathbb{E}[|X_j| \mid B] = \frac{\sqrt{3}}{2} - \frac{Bs\sqrt{3}}{48}.$$

Conditional independence gives the covariance $s^2/81$ for two distinct kernel coordinates, and the covariance $s^2/768$ for two distinct absolute coordinates. Since $\mathrm{Var}_\mu(|X|) = 1/4$,

$$\mathrm{Var}\left(\frac{1}{\sqrt{d}}\sum_j |X_j|\right) = \frac{1}{4} + \frac{(d-1)s^2}{768}.$$

The same calculation using $\mathbb{E}_\mu h^2 = 1/9$ gives (56). This proves Theorem 5.1 and the kernel-sum identity used next.

13.3 Lower bounds for conditional-variance tensorization

All conditional expectations in this proof are under $L_{d,s}$. To respect the global boundedness in Definition 5.2, replace $h(x)$ outside $[-\sqrt{3}, \sqrt{3}]$ by zero. This changes none of the following random variables or conditional integrals on the support.

First test the tensorization inequality with $f = \sum_j h(X_j)$. Conditioning on X_{-j} makes the other terms constant, so

$$\mathbb{E}\,\mathrm{Var}(f \mid X_{-j}) = \mathbb{E}\,\mathrm{Var}(h(X_j) \mid X_{-j}) \le \mathbb{E}h(X_j)^2 = \frac{1}{9}.$$

The variance of f is positive and the sum of conditional variances is nonnegative. A tensorization constant must therefore be positive, and substitution gives

$$\frac{d}{9} + \frac{d(d-1)s^2}{81} \le T\frac{d}{9}.$$

This is (53).

For the exponential bound use the bounded test $J_d = \mathbf{1}_{\{\sum_j h(X_j) > 0\}}$. Under either product component, the centered variables $h(X_j) - Bs/9$ are independent and bounded in an interval of length $10/7$. Hoeffding's exponential-moment estimate, weakened to $\mathbb{E}e^{t(h-\mathbb{E}h)} \le e^{12t^2}$, gives for a set of $m \ge 1$ coordinates and $u \ge 0$,

$$\mathbb{P}\left\{B\sum_{j=1}^{m} h(X_j) - \frac{ms}{9} \le -u \,\middle|\, B\right\} \le \exp\left(-\frac{u^2}{48m}\right). \tag{114}$$

Indeed, multiply the moment bounds and minimize $\exp(-tu + 12mt^2)$ at $t = u/(24m)$. With $m = d$ and $u = ds/9$, the error probability in identifying B by the sign of the sum is at most $e = \exp(-ds^2/3888)$ in either component. Thus $|\mathbb{E}J_d - 1/2| \le e/2$ and

$$\frac{1}{4} - \frac{1}{4}e^{-ds^2/1944} \le \mathrm{Var}(J_d) \le \frac{1}{4}.$$

This also proves (57), including its limit.

Write $R_j = \sum_{k \neq j} h(X_k)$. If $|R_j| > 1$, the value of J_d is constant as X_j varies over its support. Elsewhere its conditional variance is at most $1/4$. For $m = d - 1$ with $ms \geq 18$, $ms/9 - 1 \geq ms/18$. Applying (114) in each component therefore gives

$$\mathbb{E}\,\mathrm{Var}(J_d \mid X_{-j}) \leq \frac{1}{4}\mathbb{P}\{|R_j| \leq 1\} \leq \frac{1}{4}e^{-ms^2/15552}.$$

Under the second condition in (54), $\mathrm{Var}(J_d) \geq 1/8$. The tensorization inequality now yields $1/8 \leq Tde^{-(d-1)s^2/15552}/4$, which is (55). For every fixed $s > 0$, both conditions hold eventually. This proves Theorem 5.3.

13.4 The upper tensorization bound

We use a density comparison in which only one ratio of density bounds is lost. Let $Q = \bigotimes_{j=1}^{d} Q_j$ and $P = \rho Q$, where $0 < L \leq \rho \leq U < \infty$. For every bounded Borel f,

$$\mathrm{Var}_P f \leq \frac{U}{L} \sum_j \mathbb{E}_P \,\mathrm{Var}_P(f \mid X_{-j}). \tag{115}$$

To prove it, first use $\mathrm{Var}_P f = \inf_a \mathbb{E}_P(f - a)^2 \leq U\,\mathrm{Var}_Q f$ and product tensorization $\mathrm{Var}_Q f \leq \sum_j \mathbb{E}_Q \,\mathrm{Var}_{Q_j} f$. For each fixed x_{-j}, put $Z(x_{-j}) = \int \rho(x_{-j}, t)\, Q_j(dt)$. The conditional variational formula gives

$$Z\,\mathrm{Var}_P(f \mid x_{-j}) = \inf_a \int (f(x_{-j}, t) - a)^2 \rho(x_{-j}, t)\, Q_j(dt) \geq L\,\mathrm{Var}_{Q_j} f(x_{-j}, \cdot).$$

Integration against Q_{-j} proves $\mathbb{E}_P \,\mathrm{Var}_P(f \mid X_{-j}) \geq L\mathbb{E}_Q \,\mathrm{Var}_{Q_j} f$. Combining the inequalities proves (115). Using conditional densities and bounding each normalization separately would unnecessarily lose a second ratio; integrating the weighted conditional variance avoids that loss.

For $P = L_{d,s}$ and $Q = \mu^{\otimes d}$, each factor of either product in $\rho_{d,s}$ is at most $1 + s$. The positive-sign product alone gives the lower bound

$$L = \frac{1}{2}(1 - 3s/7)^d \leq \rho_{d,s} \leq (1 + s)^d = U.$$

In particular $L > 0$ even at $s = 1$. Substitution proves Theorem 5.4; for $d = 0$ both variances vanish.

13.5 Separation of the continuous latent law

For each sign pattern $\sigma \in \{-, +\}^d$, the map $a \mapsto (v_a(\sigma_1), \ldots, v_a(\sigma_d))$ is continuous on $[1, 2]$, and its weight is continuous and nonnegative. The finite sum of the corresponding weighted Dirac measures has mass one. Integrating against da therefore defines a probability kernel mixture: countable additivity follows directly by monotone convergence. Tonelli's theorem gives the displayed coordinate integration identity in Theorem 6.2. Direct evaluation of the two-point moments gives 0, 1, and $a^2 - 1 + a^{-2}$; integrating the last expression using the primitive $a^3/3 - a - a^{-1}$ gives $11/6$.

The separation assertion requires a null set adapted to the proposed countable mixture. Fix its factors $\mu_{r,j}$. Each probability $\mu_{r,2}$ has a countable set A_r of atoms: for every positive integer k it has at most k atoms of mass at least $1/k$. The set

$$N = \bigcup_r (A_r \cup \{-1/a : a \in A_r,\ a \neq 0\})$$

is countable. Let

$$E = \{x \in \mathbb{R}^d : x_1 \notin N, \quad x_2 = x_1 \text{ or } x_1 x_2 = -1\}.$$

This is Borel. For every $x_1 \notin N$, its section in the second coordinate consists of at most the two points x_1 and $-1/x_1$; neither is an atom of $\mu_{r,2}$. Fubini gives $(\bigotimes_j \mu_{r,j})(E) = 0$ for every r, hence $Q_d(E) = 0$.

Under P_d^{cont}, the first two coordinates always satisfy the relation defining E. Its first marginal is atomless, since both $a \mapsto a$ and $a \mapsto -1/a$ are injective and the mixing measure is Lebesgue measure. Thus $P_d^{\mathrm{cont}}(E) = 1$. By the adopted normalization of total variation, $\mathrm{TV}(P_d^{\mathrm{cont}}, Q_d) = 1$. Only this last argument requires $d \geq 2$.

A Correspondence with the formal statements

The source correspondence refers to the `ellipsoid_fitting_rocq` repository at commit `a6d542f` of its main branch. The release of the condition-number theorems at every fourth moment is tagged `phase11`, at commit `48ecb60`. The commits after the tag add the almost-everywhere lemmas of Proposition 3.4(c) and correct comments, including comments of statement files; they change no statement. The sixteen extension theorems and the original Gaussian theorem are listed separately. Supporting propositions are identified after the table. The mathematical proofs use the established inputs identified in the proof sections; the formal development also proves those inputs internally.

Mathematical content	Formal declaration and statement file
Condition-number curve at every fourth moment	`ConditionKappaCurveTheorem.condition_kappa_curve_theorem`; `extension/ConditionKappaCurveStatement.v`.
Optimal condition number at every fourth moment	`ConditionKappaTheorem.condition_kappa_theorem`; `extension/ConditionKappaStatement.v`.
Robust optimal condition number at every fourth moment	`ConditionKappaRobustTheorem.condition_kappa_robust_theorem`; `extension/ConditionKappaRobustStatement.v`.
Independent coordinates	`UniversalityTheorem.universality_theorem`; `extension/UniversalityStatement.v`.
Finite mixtures	`MixtureTheorem.mixture_theorem`; `extension/MixtureStatement.v`.
Paired laws	`PairedTheorem.paired_theorem`; `extension/PairedStatement.v`.
Latent-sign fitting	`LatentSignTheorem.latent_sign_theorem`; `extension/LatentSignStatement.v`.
Countable mixtures	`CountableTheorem.countable_theorem`; `extension/CountableMixtureStatement.v`.
Optimal condition number at fourth moment three	`ConditionTheorem.condition_theorem`; `extension/ConditionStatement.v`.
Latent-sign moment identities	`LatentScopeTheorem.latent_scope_theorem`; `extension/LatentSignScope.v`.
TV robustness and contamination	`RobustTheorem.robust_theorem`; `extension/RobustStatement.v`.
Tensorization lower bounds	`LatentTensorTheorem.latent_tensor_theorem`; `extension/LatentTensorization.v`.
Continuous latent separation	`ContinuousLatentTheorem.continuous_latent_theorem`; `extension/ContinuousLatent.v`.
Condition-curve endpoints at fourth moment three	`ConditionEndpointsTheorem.condition_endpoints_theorem`; `extension/ConditionEndpoints.v`.
Robust optimal condition number at fourth moment three	`ConditionRobustTheorem.robust_condition_theorem`; `extension/ConditionRobustStatement.v`.
Tensorization upper bound	`LatentTensorUpperTheorem.latent_tensor_upper_theorem`; `extension/LatentTensorUpper.v`.
Original Gaussian conclusions	`MainTheorem.ellipsoid_fitting_main`; `theories/MainStatement.v`.

The extension declarations have logical root `EllipsoidUniversality`; the original Gaussian declaration has root `EllipsoidFitting`. The fitting events and scalar definitions are in `theories/PaperMainDefinitions.v` and `extension/UniversalityDefinitions.v`. The other dedicated definition files, all in `extension/`, are the following:

- `MixtureDefinitions.v` and `CountableMixtureDefinitions.v`: the mixture families;

- `ConditionKappaDefinitions.v`: the curve at every fourth moment, its inverse, and the constrained variational problem;
- `ConditionDefinitions.v`: the curve at fourth moment three and the optimal data condition number `data_condition_number`;
- `RobustDefinitions.v`: point laws, total variation, samples of perturbed laws, and the contamination example;
- `PairedDefinitions.v` and `LatentSignDefinitions.v`: the dependent examples.

Four statement files also contain mathematical definitions used above:

- `extension/LatentSignScope.v` defines `latent_point_expectation`.
- `extension/LatentTensorization.v` defines coordinate replacement, conditional expectation, conditional variance, variance, and the tensorization inequality as `latent_replace`, `latent_cond_expectation`, `latent_cond_variance`, `latent_variance`, and `latent_tensorization`.
- `extension/ContinuousLatent.v` defines the coordinate values, weights, moments, and continuous point law as `cl_value`, `cl_weight`, `cl_moment`, and `cl_point_law`.
- `extension/LatentTensorUpper.v` defines the explicit upper bound `latent_tensor_constant`.

Each of the seventeen theorems has a check file, compiled by `make check`, that types the theorem against its frozen statement, prints its axioms, and compares them with an expected list. At commit `a6d542f`, the sixteen extension theorems use exactly seven axioms, those of `check/assumptions.expected` except `Classical_Prop.classic`, and the original Gaussian theorem uses all eight. They are classical axioms: propositional and functional extensionality, constructive indefinite description, Hilbert's epsilon, two decidability axioms of Stdlib's classical reals, and, for the original declaration, the additional axiom `Classical_Prop.classic`. The absence of that named dependency in the extension declarations does not make their proofs constructive.

The conjuncts of the curve statement of Theorem 3.3 are proved separately and assembled in `extension/ConditionKappaCurveTheorem.v`: the root by `condition_kappa_root_holds` in `extension/ConditionKappa/Scalar.v`; the variational value, the range, and the inverse by `condition_kappa_variational_holds`, `condition_kappa_range_holds`, and `condition_kappa_inverse_holds` in `extension/ConditionKappa/Curve.v`; the integral formula by `condition_kappa_integral_holds` in `extension/ConditionKappa/Integral.v`; and the identities at $\kappa = 3$ by `condition_kappa_three_holds` in `extension/ConditionKappa/Compatibility.v`.

The following supporting results make explicit some details that are not separate conjuncts of the seventeen main theorem declarations:

- For the curve at every fourth moment, the optimal profile, the dual form, the almost-everywhere equivalence, and the bounds of Proposition 3.4 are in `extension/ConditionKappa/Curve.v`, and the arcsine formula is in `extension/ConditionKappa/Integral.v`. The generic clipped-profile results are in `library/probability/clip_profile.v`, and the maximizer of the trace-penalized objective is in `library/matrix/trace_projection.v`. The almost-everywhere lemmas were added after the tag `phase11`.
- The Gaussian limit is `kcond_support_cvg`, proved in `extension/ConditionKappa/GaussianLimit.v`. The comparison of the diagonal and trace penalties is in `extension/ConditionKappa/Geometry.v`.

- At fourth moment three, the identities $u_{C,3} = Ct_C$ and $\ell_{C,3} = t_C$, the identity $C_\star(\alpha, 3) = C_\star(\alpha)$ for every real α, which is stronger than the frozen conjunct, the root equation `condition_t_root`, and the implications `condition_theorem_of_kappa` and `robust_condition_theorem_of_kappa` are in `extension/ConditionKappa/Compatibility.v`.
- Scalar threshold characterizations are proved in `extension/Gaussian/Threshold.v`; properties of the curve at fourth moment three and its inverse are in `extension/Condition/Curve.v`. The supporting declarations are listed below.
- The common choice of constants for finite and countable mixtures is justified by the specialization `countable_of_finite` in `extension/Countable/Theorem.v`. Preservation of those constants under robustness is a property of the proofs `robust_statement_holds` and `robust_interpolation_holds` in `extension/Countable/Robust.v`.
- Measurability of the optimal data condition number is `measurable_data_condition_number` in `extension/Condition/Theorem.v`.
- The marginal identity in every positive dimension is `latent_marginal_theorem` in `extension/LatentScopeTheorem.v`. The explicit finite-dimensional exponential tensorization bound is `latent_tensor_exp` in `extension/LatentTensorTheorem.v`.
- The countable product identity is `tuple_prod_cmixture` in `library/probability/countable_mixture.v`. Its sample-law form `rob_nu_sample` and the direct bound `robust_transfer_tv` are in `extension/Countable/Robust.v`. That file also proves the exact contamination identity as the separate lemma `contaminated_probability`.
- The continuous-law probability `cl_law`, its integral formula `integral_cl_law`, and the coordinate moments `cl_law_mean`, `cl_law_second_moment`, `cl_law_fourth_moment` are in `extension/Countable/Continuous.v`.

Sources for the data proof

The mixture covariance and variance calculation is in `extension/Features/Mixture.v` and `library/convex/feature_mixture.v`; comparison and interpolation are in `library/convex/barrier_lindeberg.v` and `extension/Mixture/Interpolation.v`. The cover, residual bounds, and threshold assembly are in `extension/Condition/Cover.v`, `extension/Condition/Slices.v`, and `extension/ConditionKappa/Sat.v` and `Unsat.v`. The measurable infimum is proved in `extension/Condition/Theorem.v`; its limit at general fourth moment is in `extension/ConditionKappa/Theorem.v`. The error transfer is in `extension/Unsat/Universality.v`. Countable and robust transfers use `extension/Countable/Truncation.v`, `extension/Countable/Robust.v`, and `extension/ConditionKappa/Robust.v`. The original prescribed-trace Gaussian input is `paper_main_conditioned_fixed_limit` in `theories/PaperMainConditionedLimit.v`.

Supporting curve declarations

For Proposition 3.4, the optimizer identities are `condition_kappa_level_gt0`, `condition_kappa_ellE`, `condition_kappa_uE`, `condition_kappa_balance`, `condition_kappa_opt_mean`, `condition_kappa_opt_value`, and `condition_kappa_opt_optimal`. They are in `ConditionKappa/Curve.v`. The scalar optimality input is `clip_optimal` in `library/probability/clip_profile.v`. The arcsine evaluation is `condition_kappa_alpha_arcsine` in `ConditionKappa/Integral.v`. The two bounds are `condition_kappa_alpha_near1` and `condition_kappa_alpha_half`. The a.e. representative argument is `clip_objective_ae_`

`set`, with the value equality `clip_value_ae`; the specialized results are `condition_kappa_value_ae` and `condition_kappa_alpha_ae`.

For Proposition 3.10, `condition_t_attained` and `condition_alphaE` identify the minimum, `condition_alpha_lip` gives the displayed Lipschitz constant, and `condition_starP` characterizes the inverse. These are in `Condition/Curve.v`. Root uniqueness and the range also follow from the general curve theorem and its compatibility identities. The total definitions use the formal conventions for empty or unbounded sets; only the stated nondegenerate domains are used here.

Quantifiers and conventions retained in this translation

1. In all mixture families, the exponential-moment bound and the first two moments hold component by component, including components of zero weight. The fourth-moment matching condition is imposed only after averaging over the mixture.

2. Constants in the mixture, paired, and latent-sign results are chosen before the families of laws and any dimension-dependent dependence parameters. An eventual dimension threshold need not be uniform over all such families.

3. Interpolation is stated outside measurable exceptional sets. No additional measurability assertion about its full event is used.

4. The frozen universality theorem states the error limit above the transition. The finite and countable mixture theorems assert it for every fixed positive density.

5. Fixed-parameter threshold and endpoint results make no joint claim with a parameter approaching a critical value as dimension grows. The condition-number results at every fourth moment are stated for fixed $\kappa > 1$, $C > 1$, and $\alpha > 0$: they make no claim at $\alpha = \alpha(C, \kappa)$, no claim uniform as $\kappa \downarrow 1$, no monotonicity claim in κ, and no endpoint expansion for $\kappa \neq 3$. The continuous latent separation theorem makes no claim about the probability of ellipsoid fitting for that law.

6. The source also formalizes the empty dimension, with empty tuples and total real operations. The matrix-fitting statements here use $d \geq 1$; the tensorization upper bound and the probability identities keep their zero-dimensional cases.

B Provenance of the proofs and scope of the contribution

The main contribution is the sharp prescribed-condition theorem for the stated mixture class, together with the calculation of its curve $\alpha(C, \kappa)$. Much of its proof architecture is inherited. This appendix identifies the inherited results, the arguments adapted to new constraints, and the particular conclusions obtained here. "New" below means additional mathematical content relative to the cited antecedents; the targeted comparison has not established publication priority. The source versions are those in the bibliography, checked on 12 October 2026.

A known theorem can have a new formal proof without becoming a new mathematical result. An extension can require a substantial argument while retaining an existing method. An elementary corollary can have useful new scope without constituting a new proof technique. The account below applies these distinctions to the dependencies of Sections 10–13.

B.1 Results and analytic inputs inherited from the literature

The unrestricted fitting problem. The Gaussian conclusions in Section 7 come from Misiakiewicz and Wen [10, Theorems 1.2–1.3, Corollary 1.4]. The fixed trace and fixed fitting level are simultaneous constraints; their conjunction is an imported theorem, not a consequence of rescaling an arbitrary fit. The row-coupling argument that passes from a fixed density to a one-sided bound on n/d^2, and the separation argument for the balanced certificate, are standard deductions.

Koehler and Sohn [8, Theorem 1.2, Proposition 1.4(ii), Theorem 1.7] supply the independent-coordinate threshold, Gaussian loss, and a general all-density loss theorem. Thus neither fourth-moment dependence of the unrestricted threshold nor convergence of the loss at every positive density is new here. The extension concerns the admissible dependent mixtures, including examples without a dimension-independent variance-tensorization bound, and the sharp constraint on the condition number. Existence of some bounded-condition fit was already included in their Theorem 1.2(i); optimizing its condition number is a further claim.

The unrestricted scalar calculation. Section 10.1 closely adapts Koehler and Sohn [8, Lemmas 5.7–5.9 and Proposition 5.10]. The cutoff equation and the unrestricted fourth-moment duality belong to that antecedent. They are retained to fix normalizations and identify the limiting range of the constrained curve. The signed dual problem for the clipped profile in Section 10.4 is a different calculation.

The probabilistic comparison machinery. Gaussian concentration, Gordon comparison, and the convex Gaussian min–max theorem are established inputs [7, 18]. The projection interpretation of Gaussian width and its connection with statistical dimension are also established [1, Theorem I, Proposition 10.2]. Their use in Section 11.5 does not introduce a new comparison principle. The witness and gap apply these tools to the evaluated cone, track its inner and outer radii, and remove the free intercept by projecting off the constant vector. Their explicit constants are quantitative bounds for this application; no sharpness is claimed for those constants.

On the data side, the direct source is Koehler and Sohn [8]: Theorem 2.9 and Section 2.2.3 for barrier comparison; Propositions 2.1 and 2.6 and Section 4.2 for approximation and feature regularity; Theorems 1.9 and 2.3 and Section 3 for interpolation; Proposition 2.4 for vector anti-concentration; and Lemmas 2.11–2.12 and Section 4.3 for regularization removal. The Lindeberg replacement, leave-one-out residual estimate, sparse and light-tail decomposition, and ridge-removal strategy in Section 12 follow this architecture. These substantial ingredients should receive credit even where the exposition states their usable form instead of repeating

their proofs. The vector anti-concentration bound used in interpolation and the small-ball localization used in ridge removal are distinct inherited inputs.

B.2 The condition cone: known geometry and the added calculation

The object and the projection method are known. Tropp [19, Sections 3.1–3.2] treats the condition cone and its dual. Bandeira and Maillard [3, Section 3.1.3] studies the same cone and its asymptotic width: Lemma 3.5 in arXiv v2, Lemma 3.6 in the published version. The curve at fourth moment three evaluates their width function, with the normalization $f(C) = 2\sqrt{\alpha_C}$. It is not the introduction of a previously unknown cone or Gaussian width problem.

Clipping eigenvalues to $[t, Ct]$ and optimizing the scalar level is established condition-constrained matrix approximation. The primary abstract of Tanaka and Nakata [17] confirms the reduction for unitarily invariant norms; we do not assign a theorem number from its unavailable full text. Won et al. [20, Theorem 1, Section 2.2] gives a related clipping optimization for a covariance likelihood, a different objective. Spectral alignment, convex projection, and scalar first-order optimality are therefore not claimed as original methods.

What is calculated here. Sections 10.2–10.5 solve the condition-constrained semicircle problem for every $C > 1$ and $\kappa > 1$. The particular addition is the coupled pair of clipping points, its unique root, and the identification of the optimized value with $\alpha(C, \kappa)$. The root argument does not assume monotonicity of the auxiliary map $L_{C,\kappa}$. The optimality remainder stays coercive for negative $c = (\kappa - 3)/2$, since $1 + c > 0$. The signed scalar dual has a minimum for $c > 0$ and a maximum for $c < 0$; using one direction for both signs would not prove the stated range. These are specific derivations for the constrained problem using ordinary convex and quadratic optimization.

The integral and arcsine formulas, strict monotonicity, full range, inverse, and fourth-moment-three compatibility complete that characterization. They are consequences of the optimized profile, proved by integration, perturbation, and continuity. Replacing an almost-everywhere admissible profile by its pointwise clipped version is likewise an elementary representative argument. None of these methods is a separate originality claim.

What the endpoint coefficient establishes. The lower-clipping trial of Bandeira and Maillard [3] and the positive-part-plus-constant trial of Misiakiewicz and Wen [10, Lemma 2.2] already yield the leading coefficient $8/(3\pi)$ after expansion; the calculation is displayed in Section 8. The papers do not state that expansion, but its availability from their trials prevents a claim that the coefficient itself is new. The additional conclusion is the matching obstruction, the exact curve at all C, and its transfer to exact data fits. Section 10.7 also derives the small-density expansion and controlled remainders from the curve. Inverting the curve gives the asymptotic of $C_\star$; inversion is not an independent method.

The replica prediction of Maillard and Kunisky [9, Sections 1.3.1 and 3.3] is a conceptual antecedent: it optimizes a lower eigenvalue constraint. It is one-sided, so it neither imposes the ratio constraint nor supplies the rigorous theorem proved here.

B.3 The fourth-moment deformation: a substantive adaptation

The antecedent is more than the semicircle law. The map T_κ, Haar diagonal control, and diagonal-to-trace reduction already occur in Koehler and Sohn [8, Section 5.1, Lemmas 5.13–5.14, Proposition 5.15]. The present extension substitutes the condition cone $T_\kappa \mathcal{K}_C$ for the transformed positive semidefinite cone. The covariance map is inherited; evaluating the new constrained support function is the problem to be solved.

The added comparison argument. Sections 11.2–11.3 maximize the trace-penalized quadratic on $\mathcal{K}_C$. Uniqueness gives orthogonal equivariance of the maximizer $X(H)$; the center of its weighted projection is $\mathcal{A}_c^{-1}H$. Rotation invariance controls the diagonal fluctuation of $X(H)$, while strong concavity bounds the possible improvement of the original objective away from $X(H)$. Both estimates are required to compare the two suprema. The second estimate makes the argument work directly for $-1 < c < 0$ as well as $c \geq 0$, without treating the condition cone as if it were the positive semidefinite cone and had its duality properties.

This adapts the existing reduction, with an additional argument needed by the condition constraint. The elementary rotation identity itself is not new: the proof here obtains it from a two-coordinate rotation and uses an integrable equivariant maximizer, rather than conditioning on a bounded spectral profile. This is an expository variant, not a claim of priority for the identity.

The shifted limit. Section 11.4 combines the trace penalty with the known condition-cone projection. Its finite-dimensional expression retains the random trace term until a uniform estimate removes it. The shift and clipping level both vary, so pointwise semicircle convergence would not justify substituting the random optimizing shift. A compact two-parameter net supplies the needed uniformity, and the signed scalar identity identifies the limit. The particular constrained limit is an addition here; finite nets and equicontinuity are standard tools. This separates the new limit from the familiar method used to exchange optimization and convergence.

B.4 Mixtures and the sharp threshold for data

Mixture regularity is an additional verification. Section 12.1 conditions on the latent label for the product estimates, then uses total variance. The covariance of a component need not equal the averaged covariance. Its fluctuation contributes the additional squared covariance-spread term in Lemma 12.1. Controlling that term proves feature regularity for the full mixture without imposing dimension-independent tensorization on the point law. Applying a common-fourth-moment product theorem separately to each component would miss this issue: only the averaged fourth moment is fixed. This verification extends the class of point laws covered by the fitting conclusions; the comparison theorem is inherited.

Conditional interpolation closely follows the existing proof. Section 12.2 needs estimates uniform over independent arrays whose coordinate laws can differ between rows. The formal development tracks this uniformity in the existing sparse and light-tail argument. Once the nuclear-norm bound is available, allowing every residual follows by convex separation. Removing a range restriction is thus a consequence of the bound, not a new interpolation principle. For mixtures, conditioning on the entire tuple of labels and intersecting the exceptional sets transfers that uniform estimate without a factor depending on the number of components. This is the specific assembly needed to permit an unbounded number of labels.

The prescribed-condition assembly is the main data addition. Sections 12.3–12.4 preserve a strict spectral margin while applying the inherited approximation and correction machinery. A finite collection of spectral windows covers the normalized condition cone. Simultaneous concentration permits a deterministic choice of window; choosing a random window first would not directly justify the expectation comparison. Below the curve, the small operator-norm correction fits all points while staying inside the prescribed condition bound. Above the curve, the Gaussian gap and control of the free intercept exclude every fit under that bound. The two directions together, for the same explicitly evaluated curve, are stronger than existence of some bounded-condition fit.

Compact-projection measurability, the second-moment bound on the intercept, and bracketing the infimum by two fixed condition bounds complete this theorem and the limit of the optimum. These are necessary technical steps using standard methods, not separate claims of mathematical originality. No attainment assumption is hidden in the argument for the optimal data condition number.

B.5 Transfers and examples: new conclusions from standard methods

Countable mixtures and robustness. Truncation and coupling are standard. The important detail in Section 12.7 is to replace the discarded tail by one product of its coordinate-mixture marginals. This preserves the averaged fourth moment exactly, as well as the common subgaussian bound. Deleting and renormalizing the tail would not do so. Combined with constants independent of the number of finite components, this gives the countable result without a uniform tail-decay assumption.

The product estimate $\mathrm{TV}(P^{\otimes n}, Q^{\otimes n}) \le n\,\mathrm{TV}(P, Q)$ is elementary. The robustness conclusions apply it to the proved measurable fitting events and exceptional sets; they introduce no new total-variation principle. The absence of moment assumptions on the perturbed law follows from this probability-level transfer. Zero-point contamination and the limit $(1 - \lambda/n)^n \to e^{-\lambda}$ identify an obstruction at the boundary scale. This shows why $o(d^{-2})$ cannot be replaced by $O(d^{-2})$ in a uniform inheritance theorem. It does not assert that every particular larger perturbation destroys fitting.

Paired and latent-sign laws. The paired density belongs to the classical family of Sarmanov [13]. Its fourth-Legendre specialization, the full interval $[-1, 7/3]$, and the two product-mixture decompositions in Section 13 are explicit calculations. Their purpose is to put the example inside the fitting theorem with constants valid at both endpoints. They are not a new construction of dependent density families. Similarly, conditioning on the latent sign gives the displayed covariance identities by standard total-covariance calculations. The particular calibration preserves the required marginal moments while making persistent dependence explicit.

Tensorization bounds. The conditional-variance energy is classical: it is the Dirichlet form of single-coordinate heat-bath updates [4, Section 1.1, equations (1.5)–(1.6)]. General deterioration of functional inequalities under mixing also predates this example [16]. The additive test for the linear lower bound, the bounded component-detecting test for the exponential lower bound, Hoeffding's inequality, and comparison with a product density are standard mechanisms. The contribution is their explicit application to the same law that satisfies the fitting theorem. It gives the stated linear bound, the constants $1/15552$ and $1/2$ in the exponential lower bound, and the upper bound $2((1+s)/(1-3s/7))^d$ throughout $0 \le s \le 1$. The upper proof compares integrated conditional variances and retains one density ratio. The generic comparison is not claimed as new, and the exponential rates are not claimed to be optimal.

The continuous latent obstruction. The countable-atoms and Fubini argument is a standard singular-support argument; even an atomless diagonal law illustrates its mechanism. Here the two-point conditional law on $\{a, -1/a\}$ additionally has mean zero, variance one, and bounded support. Its continuous mixture retains these calibrations, yet is at total-variation distance one from every countable product mixture in dimension at least two. The null set is chosen after the candidate countable mixture, removing all relevant coordinate atoms. This handles arbitrary competing product factors, including atomic ones. The calibrated example

is the addition; the general singularity mechanism is classical, and the result makes no claim about the probability of ellipsoid fitting for this law.

B.6 What is proved here, cited here, and formally checked

The mathematical exposition proves the constrained scalar calculation, the condition-cone comparison, the mixture application, and the transfers and examples. It states the barrier comparison, nuclear-norm estimate, unrestricted Gaussian-loss result, and prescribed-trace Gaussian theorem as analytic inputs. Their complete antecedent proofs are not reproduced in this document. Appendix A identifies the corresponding internal formal proofs: the Rocq development also proves those inputs, using its listed classical axioms.

Thus the principal candidate for an original mathematical contribution is the exact curve and sharp optimal-condition theorem for the stated finite and countable mixture class, together with the adaptations above that establish it. The supporting identities, examples, and robustness results have the individual status described above. Neither formal verification nor an independently written proof establishes publication priority; that remains a qualified literature assessment.

C The reusable Rocq libraries

The development separates general mathematics from the ellipsoid-fitting theorems. At commit `a6d542f`, `library/` contains 82 Rocq modules under the logical root `GaussianAnalysis`. These modules construct probability laws and prove analytic, geometric, and optimization results with explicit hypotheses independent of the point model. The fitting statements and their applications reside in `theories/`, under `EllipsoidFitting`, and `extension/`, under `EllipsoidUniversality`. This appendix describes the reusable interfaces and their role in the proof. Appendix B separately addresses mathematical provenance: formalizing an established inequality in a new library is not a claim that the inequality is new.

C.1 Foundations, organization, and conventions

The library builds on MathComp's algebra and finite-dimensional matrix types and MathComp-Analysis's topology, measures, integration, and one-dimensional normal law. Its analytic statements are parameterized by an abstract `R:realType`; algebraic lemmas use weaker structures when sufficient. The paper instantiates these results at Stdlib's real numbers through bridge modules outside the library. The code is organized for reuse within this project; no upstream incorporation into MathComp-Analysis is asserted.

Two conventions are essential when using the interfaces. A matrix's ambient MathComp norm is the largest absolute entry. Euclidean and Frobenius geometry instead use `mxdot` and `frobenius`, and the operator norm is `opnorm`. Substituting the ambient norm in a Gaussian or spectral bound would change its dimension dependence. Also, probabilities and nonnegative integrals often take values in the extended reals. A real-valued expectation formula requires the stated integrability hypotheses; an integral expression alone is not a finiteness certificate.

The following counts are the source modules registered in `_CoqProject`, rather than a count of individual declarations. The seven directories group topics; they are not seven independent packages or a strict hierarchy of dependencies.

Directory and modules	Reusable mathematics and its use here
`gaussian/` (9)	Measurable matrices, finite-dimensional Gaussian and Lebesgue measures, moments, Fubini, integration by parts, and covariance interpolation. This supplies the measure and calculus foundation for Gaussian comparison.
`comparison/` (19)	Gaussian linear images, rotation and translation, concentration, Slepian and Gordon inequalities, Sudakov bounds, cone escape, a convex Gaussian min–max specialization, and the Ornstein–Uhlenbeck and Brascamp–Lieb machinery. These yield witnesses, gaps, and the analytic input to small-ball estimates.
`matrix/` (13)	Real symmetric spectral theory, Frobenius, operator and nuclear norms, nets, coverings, whitening, positive parts, condition cones, and trace-penalized projection. These connect spectral constraints to Gaussian geometry and residual interpolation.
`probability/` (19)	Product laws, conditioning, finite and countable mixtures, densities, atoms, total variation, moment and tail bounds, Efron–Stein, Hanson–Wright, symmetrization, measurable minimizers, and scalar variational problems. These support both dependent laws and optimization with random data.
`fourier/` (6)	Characteristic functions, Gaussian regularization, Esseen bounds, concentration of sums, vector small-ball estimates, and Fourier inversion. The small-ball estimate is used in exact interpolation.

Directory and modules	Reusable mathematics and its use here
`convex/` (12)	Self-concordance, log-determinant barriers, regularized least squares, minimizers, insertion costs, feature moments, mixture regularity, Lindeberg comparison, barrier removal, ridge comparison, and Moreau projection. These transfer Gaussian optimization to data.
`random_matrix/` (4)	GOE moments and rotation invariance, semicircle limits, and restricted-isometry estimates for independent subexponential columns. These supply the spectral limit and uniform data estimates.

C.2 Gaussian analysis, comparison, and small-ball estimates

A finite-dimensional Gaussian probability. In `gaussian/gauss.v`, `gauss` is the standard Gaussian probability on MathComp row vectors. The characterization `gauss_unique`, coordinate Fubini, and block-product identities make independence available as measure identities. The matrix measurability interface `measurable_fun_mxP` reduces measurability to entries. The calculus modules prove `gauss_ibp` under derivative, continuity, and polynomial-growth hypotheses, and `gauss_interpolation` for smooth functions along a differentiable matrix path with the stated growth and boundedness controls. These are reusable analytic theorems, rather than formal assumptions standing in for integration by parts or differentiation under an integral.

The precise comparison interfaces. `gauss_mulmx_law` identifies Gaussian linear images with the same covariance, even when their input dimensions differ. `gauss_concentration` gives the Gaussian exponential-moment bound for globally Lipschitz functions in Euclidean norm. `gauss_poincare_lipschitz` gives its variance bound; this exported result should not be confused with the more general gradient-energy Poincaré inequality.

The module `comparison/cgmt.v` implements the comparison for $\inf_{x\in B}\|Ax\|$ used in Section 11.5. For a nonempty set B bounded by radius ρ and a positive row dimension, `cgmt_lower` compares means with additive error ρ; `cgmt_upper` additionally requires B closed and convex. Concentration and the interfaces `cone_escape_zero_tail` and `cone_escape_lower_tail` turn this into witness and separation bounds for cone sections. This is the implemented specialization, not an unrestricted arbitrary-objective min–max API.

Fourier analysis and anti-concentration. The characteristic function is represented by its cosine and sine integrals, with multiplication implemented on pairs of reals. The Brascamp–Lieb modules prove the geometric rank-one inequality through the Mehler semigroup, first for bounded positive functions and then for nonnegative measurable functions. Combined with Gaussian regularization and whitening, this yields `vector_small_ball`. It allows independent, nonidentically distributed scalar laws with common second-moment and anti-concentration bounds. If the coefficient vectors satisfy $\sum_i w_i w_i^\mathsf{T} \succeq I_m$ and $\|w_i\| \le \rho$, it bounds $\Pr(\|\sum_i X_i w_i - z\| \le \eta\sqrt{m})$ by $(C\eta)^m$, uniformly in z, provided $\eta \ge C\rho$. The constant depends only on the common moment and anti-concentration parameters. The moment-based specialization `vector_small_ball_moments` is used by `extension/Sat/OneStage.v`. If the characteristic-function modulus is integrable, `cf_rV_inversion` identifies a continuous bounded density by Fourier inversion. The small-ball proof does not require this result.

C.3 Matrix geometry and the spectral limit

Condition cones and deterministic interpolation. The spectral and norm modules provide real symmetric diagonalization, Hoffman–Wielandt estimates, and operator/nuclear norm duality. In `matrix/condition_cone.v`, `cond_cone_dist` reduces the distance of a

symmetric matrix to $\mathcal{K}_C$, for $C \geq 1$, to scalar eigenvalue clipping. `cond_support_sqr` relates this distance to Gaussian support, and `cond_slice_frobenius` gives the radius bounds for its trace slice. These are deterministic geometric statements; their random spectral limit is proved separately. On the data side, `range_duality_sym` in `matrix/nuclear.v` converts a nuclear norm lower bound into a symmetric interpolating matrix with a controlled operator norm. The random-matrix and covering modules establish the estimates to which that duality is applied.

A general trace-penalized maximizer. `matrix/trace_projection.v` works on any nonempty closed convex set K of square matrices, not just symmetric matrices or condition cones. For $n \geq 1$ and $c > -1$, define

$$\Phi_c(H, S) = 2\langle H, S\rangle - \|S\|_F^2 - \frac{c}{n}(\operatorname{tr} S)^2, \qquad a_c = \min\{1, 1 + c\}.$$

The definition `trace_argmax` gives its unique maximizer $X(H)$. Besides the variational inequality `trace_argmaxP`, the interfaces `trace_argmax_ge` and `trace_argmax_lipschitz` give, for $S \in K$,

$$\Phi_c(H, S) + a_c\|S - X(H)\|_F^2 \leq \Phi_c(H, X(H)), \qquad \|X(H) - X(H')\|_F \leq a_c^{-1}\|H - H'\|_F.$$

The construction is a weighted projection centered at $\mathcal{A}_c^{-1}H$, where $\mathcal{A}_c S = S + (c/n)(\operatorname{tr} S)I_n$. If $U_c = \mathcal{A}_c^{1/2}$, it is $U_c^{-1}\Pi_{U_c K}(U_c^{-1}H)$. Cone identities and orthogonal equivariance are separate results with the corresponding extra hypotheses, including `trace_argmax_conj`. Coercivity for $-1 < c < 0$ is essential to the application at $1 < \kappa < 3$.

Rotation and semicircle interfaces. `random_matrix/goe_rotation.v` proves GOE rotation invariance and diagonal-fluctuation estimates for measurable equivariant matrix functions that preserve symmetry. `goe_diag_var_identity` assumes entrywise square integrability; `goe_diag_var_tail` uses truncation to avoid that assumption in its tail bound. The proof uses Gaussian invariance and elementary rotations, without constructing Haar measure.

There is a normalization change at the application boundary. For a matrix A with independent standard normal entries, the library's `goe_mx` is

$$W_d = \frac{A + A^{\mathsf{T}}}{\sqrt{2d}}, \qquad H_d = \frac{W_d}{\sqrt{2}}.$$

The limiting spectral laws of W_d and the paper's H_d have radii 2 and $\sqrt{2}$, respectively. The module `random_matrix/wigner.v` proves the moment and spectral-average limits. Its `wigner_uniform` and `wigner_uniform_inf` handle compact one-parameter families of continuous profiles, uniformly Lipschitz on $[-3, 3]$ and in the parameter, with measurable exceptional sets. The additional shift parameter needed here is handled in `extension/ConditionKappa/GaussianLimit.v` by `shift_inf_unif`; that two-parameter result is not an export of `wigner.v`.

C.4 Mixtures, scalar optimization, and transfer to data

Probability constructors and exact sample laws. The files `probability/mixture.v` and `probability/countable_mixture.v` define mixtures on an arbitrary measurable type. They prove weighted integration, total variance for finite mixtures, and exact sample expansions `tuple_prod_mixture` and `tuple_prod_cmixture`. For countable mixtures, nonnegative integration needs measurability; signed integration additionally requires integrability under every component and summability of the weighted absolute integrals. The finite result

`mixture_cover` transfers measurable covers of a common bad set by intersection, without a union bound over labels. The bad set itself need not be measurable.

The interface `tuple_prod_tv` in `probability/total_variation.v` bounds the discrepancy of product laws by the sum of their coordinate total variations. It permits different laws in every coordinate and needs no moments or densities. The density constructors and countable-atom lemmas support the explicit dependent examples. For reuse, `density_law` takes a proof that the density is valid; the totalized constructor `density_prob` falls back to its reference law when that condition fails.

Regularity of mixed features. The definition `regular_feature` combines centering, covariance, a directional fourth-moment bound, a variance bound for quadratic forms, and positivity of the covariance operator norm. `mixture_regular_feature_dot`, in `convex/feature_mixture.v`, is a finite-mixture closure theorem for this class. Let centered components have covariance Σ_i, common fourth-moment bound $A\|\theta\|^4$, and common quadratic-form variance bound $\tau^2\|T\|_F^2$. With $\Sigma = \sum_i w_i\Sigma_i$, assume

$$\sum_i w_i\langle\Sigma_i - \Sigma, T\rangle^2 \le b^2\|T\|_F^2 \quad (T = T^{\mathsf{T}}), \qquad \|\Sigma\|_{\mathrm{op}} > 0, \quad A \le K^4\|\Sigma\|_{\mathrm{op}}^2.$$

The mixture is then $(K, \sqrt{\tau^2 + b^2})$-regular. `mixture_regular_feature` gives the sufficient condition $\sum_i w_i\|\Sigma_i - \Sigma\|_F^2 \le b^2$. These statements are not restricted to quadratic data features or product components. Section 12.1 verifies their hypotheses for the point laws considered here.

A scalar optimizer for general laws. The files `probability/variational.v` and `probability/clip_profile.v` separate scalar optimization from semicircle integration. For a symmetric law with finite second moment, $C \ge 1$, $t \ge 0$, and $c > -1$, the clipped profile $q(x) = \min\{\max\{x - a, t\}, Ct\}$ is optimal for $2\mathbb{E}[Xq(X)] - \mathbb{E}[q(X)^2] - c(\mathbb{E}q(X))^2$ when its balance and fixed-point conditions hold. `clip_optimal` proves global optimality, and `clip_dual_le`, `clip_dual_ge`, and `clip_dual_opt` give the dual comparisons with the correct signs. The generic theorem assumes these scalar conditions; it does not solve the semicircle root equation. That solution and the arcsine formula belong to `extension/ConditionKappa/`.

The formal equality `clip_value_ae` identifies the pointwise and almost-everywhere constrained suprema for every probability law and all real c, C. It does not require symmetry or a second moment. The separate lemma `clip_admissible_ae_integrable` proves finiteness of all three integrals under a finite-second-moment assumption. This separates an elementary representative argument from the analytic assumptions of the optimizer theorem.

Measurable optimization and barrier comparison. `probability/measurable_minimizer.v` proves measurability of a minimizer from quadratic growth, a fixed countable sequence approximating its value, and measurability of the objective along that sequence. Its least-squares exports `measurable_ls_min` and `ls_min_tuple_set` give measurability and invariance under changes to an unused row. Together with independent rows, that invariance justifies the leave-one-out argument. These are mathematical selections, not certified numerical algorithms.

The convex modules isolate the deterministic analysis of self-concordance, log-determinant barriers, minimizers, and insertion costs from the probabilistic replacement step. `barrier_lindeberg_mean`, `barrier_lindeberg_variance`, and `barrier_lindeberg_concentration` compare regularized least-squares values for two regular feature laws with the same covariance. The final interfaces use independent identically distributed rows under each law, a nonempty relatively open convex domain, a nonnegative `sc_function`, a nonnegative regularity parameter

τ, and positive ridge and barrier parameters. Both the mean comparison and concentration bound require $n\gamma \geq 36$; the variance bound does not.

One hypothesis deserves care: `sc_function` includes divergence at the boundary along finite exit segments. The comparison permits an unbounded domain and dispenses with the additional gradient-versus-Hessian parameter bound of `sc_barrier`; it does not dispense with boundary blow-up. Barrier removal is a library result; the ridge-removal step is `unsat_ridge_removal` in `extension/Unsat/Ridge.v`. Section 12.2 gives the quantitative comparison used in the fitting proof.

C.5 Application boundaries and reuse

The generic library ends before the model-specific assembly. The covariance deformation of the condition cone and its spectral limit, the finite cover of condition-preserving spectral slices, and the sharp data theorem are in `extension/ConditionKappa/` and `extension/Condition/`. Likewise, the exact product-mixture identities are generic, but replacement of a discarded label tail by a product preserving its coordinate moments is proved in `extension/Countable/Truncation.v`. The explicit paired, latent-sign, and continuous-latent laws remain application modules. This distinction prevents an application theorem from being mistaken for a library interface available under weaker hypotheses.

For example, the trace-projection module has logical name `GaussianAnalysis.matrix.trace_projection`. With the project's load path and compiled dependencies, representative interfaces can be inspected with:

```
From GaussianAnalysis.matrix Require Import trace_projection.
From GaussianAnalysis.probability Require Import clip_profile.
Check trace_argmaxP.
Check clip_optimal.
Print Assumptions clip_optimal.
```

The source header of each module states its definitions and main results. `library/README.md` provides the larger declaration index; the actual theorem and its section hypotheses determine the contract. The documented toolchain is Rocq 9.1.1, MathComp 2.6.0, and MathComp-Analysis 1.16.0, with Hierarchy Builder 1.10.3.

The checker rejects library imports from either paper namespace, Stdlib's real-analysis modules, or the MathComp bridges `Rstruct` and `Rstruct_topology`. It also rejects unfinished proofs and new axiom declarations in the checked sources. This structural independence does not mean axiom-freedom: the development uses classical foundations, and a reused lemma's actual dependencies are those reported by `PrintAssumptions`. Appendix A records the reports for the main theorems; those counts are not per-module axiom counts. Kernel checking establishes the formal implication, while mathematical review must still verify that the interface states the intended hypotheses and conclusion.

References

[1] Dennis Amelunxen, Martin Lotz, Michael B. McCoy, and Joel A. Tropp. Living on the edge: Phase transitions in convex programs with random data. *Information and Inference*, 3(3):224–294, 2014. doi: 10.1093/imaiai/iau005. URL https://arxiv.org/abs/1303.6672.

[2] Aykut Arslan. The strict threshold for Gaussian ellipsoid fitting, 2026. URL https://ai.meta.com/research/publications/the-strict-threshold-for-gaussian-ellipsoid-fitting/. Research paper, 2 October 2026.

[3] Afonso S. Bandeira and Antoine Maillard. Exact threshold for approximate ellipsoid fitting of random points. *Electronic Journal of Probability*, 30:1–46, 2025. doi: 10.1214/25-EJP1378. URL https://arxiv.org/abs/2310.05787. Paper no. 126.

[4] Pietro Caputo, Georg Menz, and Prasad Tetali. Approximate tensorization of entropy at high temperature. *Annales de la Faculté des Sciences de Toulouse. Mathématiques*, 24(4): 691–716, 2015. doi: 10.5802/afst.1460. URL https://www.numdam.org/item/AFST_2015_6_24_4_691_0/.

[5] Yatin Dandi, Ludovic Stephan, Florent Krzakala, Bruno Loureiro, and Lenka Zdeborová. Universality laws for Gaussian mixtures in generalized linear models. In *Advances in Neural Information Processing Systems*, volume 36, pages 54754–54768, 2023. URL https://arxiv.org/abs/2302.08933.

[6] Sofia de la Cerda, Aaron Potechin, Madhur Tulsiani, and Jeff Xu. Sharp phase transition for ellipsoid fitting, 2026. URL https://arxiv.org/abs/2608.12415v1. arXiv:2608.12415v1.

[7] Y. Gordon. On Milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$. In Joram Lindenstrauss and Vitali D. Milman, editors, *Geometric Aspects of Functional Analysis*, volume 1317 of *Lecture Notes in Mathematics*, pages 84–106. Springer, 1988. doi: 10.1007/BFb0081737. URL https://doi.org/10.1007/BFb0081737.

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

[9] Antoine Maillard and Dmitriy Kunisky. Fitting an ellipsoid to random points: Predictions using the replica method. *IEEE Transactions on Information Theory*, 70(10): 7273–7296, 2024. doi: 10.1109/TIT.2024.3415654. URL https://arxiv.org/abs/2310.01169v2.

[10] Theodor Misiakiewicz and Garrett G. Wen. The sharp SAT/UNSAT phase transition in random ellipsoid fitting, 2026. URL https://arxiv.org/abs/2608.10184v1. arXiv:2608.10184v1.

[11] Andrea Montanari and Basil N. Saeed. Universality of empirical risk minimization. In *Proceedings of the 35th Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pages 4310–4312, 2022. URL https://arxiv.org/abs/2202.08832. Full paper: arXiv:2202.08832.

[12] Mark Rudelson and Roman Vershynin. Small ball probabilities for linear images of high-dimensional distributions. *International Mathematics Research Notices*, 2015(19): 9594–9617, 2015. doi: 10.1093/imrn/rnu243. URL https://arxiv.org/abs/1402.4492.

[13] O. V. Sarmanov. Generalized normal correlation and two-dimensional Fréchet classes. *Doklady Akademii Nauk SSSR*, 168(1):32–35, 1966. URL https://www.mathnet.ru/eng/dan32257.

[14] James Saunderson, Venkat Chandrasekaran, Pablo A. Parrilo, and Alan S. Willsky. Diagonal and low-rank matrix decompositions, correlation matrices, and ellipsoid fitting. *SIAM Journal on Matrix Analysis and Applications*, 33(4):1395–1416, 2012. doi: 10.1137/120872516. URL https://arxiv.org/abs/1204.1220.

[15] James Saunderson, Pablo A. Parrilo, and Alan S. Willsky. Diagonal and low-rank decompositions and fitting ellipsoids to random points. In *Proceedings of the 52nd IEEE Conference on Decision and Control*, pages 6031–6036, 2013. doi: 10.1109/CDC.2013.6760842. URL https://dspace.mit.edu/handle/1721.1/90825.

[16] André Schlichting. Poincaré and log-Sobolev inequalities for mixtures. *Entropy*, 21(1):89, 2019. doi: 10.3390/e21010089. URL https://arxiv.org/abs/1812.06464.

[17] Mirai Tanaka and Kazuhide Nakata. Positive definite matrix approximation with condition number constraint. *Optimization Letters*, 8(3):939–947, 2014. doi: 10.1007/s11590-013-0632-7. URL https://doi.org/10.1007/s11590-013-0632-7.

[18] Christos Thrampoulidis, Samet Oymak, and Babak Hassibi. Regularized linear regression: A precise analysis of the estimation error. In *Proceedings of the 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pages 1683–1709, 2015. URL https://proceedings.mlr.press/v40/Thrampoulidis15.html.

[19] Joel A. Tropp. Integer factorization of a positive-definite matrix. *SIAM Journal on Discrete Mathematics*, 29(4):1783–1791, 2015. doi: 10.1137/15M1024718. URL https://arxiv.org/abs/1506.00340v3.

[20] Joong-Ho Won, Johan Lim, Seung-Jean Kim, and Bala Rajaratnam. Condition-number-regularized covariance estimation. *Journal of the Royal Statistical Society: Series B*, 75(3):427–450, 2013. doi: 10.1111/j.1467-9868.2012.01049.x. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3667751/.