ASYMPTOTIC ANALYSIS OF EMPIRICAL RISK MINIMIZATION ON ENTRY-WISE I.I.D. HEAVY-TAILED DATA

Anonymous authors
Paper under double-blind review

ABSTRACT

Many real-world datasets exhibit unusually large values far more frequently than predicted by Gaussian models. Heavy-tailed distributions capture this behavior, yet evaluating learning performance under them remains challenging because rare, large feature entries retain non-vanishing effects even in high dimensions. Even in the canonical setting of empirical risk minimization for linear regression with entry-wise i.i.d. symmetric α-stable data, a precise asymptotic characterization of prediction has been lacking. In this work, we introduce a functional order parameter that describes the random effective problem associated with each coefficient. Using the replica method, we fully characterize the generalization error in the proportional high-dimensional limit where the sample size and feature dimension diverge at a fixed ratio. Additionally, this analysis establishes a heavy-tail universality law, scaling laws relating typical errors to prediction reliability, and the Bayes-optimal prediction error. In addition to characterizing the effects of extreme entries on the learning process, our method applies broadly to other systems with persistent local heterogeneity.

1 INTRODUCTION

Real-world data in finance (Mandelbrot, 1963), communications (Ilow & Hatzinakos, 1998; Middleton, 1977), natural images (Wainwright & Simoncelli, 1999), and the life sciences (Furusawa & Kaneko, 2003) often contain outliers and extreme observations much more frequently than predicted by a Gaussian distribution. When even the feature variance does not exist, the Gaussian universality suggested by the usual central limit theorem may no longer hold, and tail behavior can have a non-vanishing effect on learning even in the high-dimensional limit. Heavy-tailedness therefore not only poses a robustness issue, but can fundamentally change the statistical theory of high-dimensional learning.

The theory of learning from heavy-tailed data has mainly developed in two directions. The first is non-asymptotic statistics, which provides error guarantees under weak moment assumptions (Brownlees et al., 2015; Mendelson, 2014; 2018; Lecué & Mendelson, 2018; Lecué & Lerasle, 2020). These results apply to a broad class of data distributions, including models with independent heavy-tailed coordinates. However, they mainly provide worst-case error bounds and do not precisely characterize typical high-dimensional errors or their dependence on the loss, regularization strength, and tail index.

The second direction studies typical high-dimensional behavior in the proportional limit, where the dimension D and sample size N grow at the same rate. For heavy-tailed features, existing precise results rely almost exclusively on elliptical models $\boldsymbol{x}_\mu = \sigma_\mu \boldsymbol{z}_\mu$ with $\boldsymbol{z}_\mu \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}/D)$, where a single random variable σ_μ scales all coordinates of sample μ (Adomaityte et al., 2023; 2024). However, this reliance on elliptical symmetry has two major limitations: First, from a theoretical standpoint, prior elliptical models introduce heavy-tailedness through a single random scale σ_μ per sample. Conditioning on this scale restores a Gaussian feature vector, making these models a more restricted class of high-dimensional heavy-tailed distributions. In contrast, when coordinates are independently heavy-tailed, extreme values appear in isolated entries, so existing analytical techniques inapplicable. Second, in high-dimensional sensing and measurement systems, including communication, signal detection, and imaging applications, localized interference and physical events can generate extreme observations concentrated in a small number of coordinates (Karakuş et al., 2020;

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

Chopra & Evans, 2012; Shen & Qin, 2016). For such applications, an entry-wise heavy-tailed model is far more realistic than an elliptical one that scales all features through a shared random factor.

To bridge this gap, we first study empirical risk minimization (ERM) with coordinate-wise i.i.d. symmetric α-stable features ($0 < \alpha < 2$) in the proportional high-dimensional limit, where $N, D \to \infty$ and $N/D \to \kappa \in (0, \infty)$. The corresponding Gaussian-feature problem is a canonical setting in high-dimensional statistics, and its typical performance has been studied for many years (Sompolinsky et al., 1990; Seung et al., 1992; Watkin et al., 1993; El Karoui et al., 2013; Donoho & Montanari, 2016; Thrampoulidis et al., 2018). Yet, after replacing Gaussian features with independent heavy-tailed features, no general asymptotic theory is available for this canonical problem. We therefore ask how the typical performance of the learned estimator can be characterized in this setting.

We use the replica method (Mézard et al., 1987; Charbonneau et al., 2023) from statistical physics to reduce the problem to an effective one-dimensional optimization. For Gaussian features, as in many standard replica calculations, this effective problem is described by a finite set of scalar order parameters. By contrast, for coordinate-wise independent heavy-tailed features, a single extreme entry can preserve the full nonlinear response of the loss in the effective problem. The samples producing such extremes differ across coordinates, and this heterogeneity persists in high dimensions. The system therefore requires a *functional* order parameter $\hat{\mathcal{P}}$, defined as the distribution of these random response functions. We derive a self-consistent equation for this functional order parameter and use it to compute the learned coefficient distribution and typical estimation error.

Our main results are as follows.

- Section 4 characterizes the high-dimensional ERM through a functional order parameter, reducing it to a random one-dimensional effective optimization problem and determining its asymptotic generalization error.
- Section 5.1 establishes heavy-tail universality: the learned behavior depends on the feature distribution only through its tail index and total tail constant.
- Section 5.2 shows that typical prediction errors decrease as $\kappa^{-1/\alpha}$, faster than the Gaussian $\kappa^{-1/2}$ rate, while the probability of a fixed large error decreases only as κ^{-1} rather than exponentially.
- Section 5.3 characterizes Bayes-optimal prediction performance and shows that entry-wise heavy-tailed features outperform their elliptical counterparts at finite κ, although both exhibit the same leading-order scaling for $\kappa \gg 1$.

2 Related Work

Analysis with functional order parameters. Heterogeneous systems with sparse or heavy-tailed interactions remain locally nonuniform even in the high-dimensional limit, and their fluctuations cannot generally be described by finitely many scalar order parameters (Mézard & Parisi, 2001; Franz et al., 2001; Neri et al., 2010; Jung, 2018). Discrete-variable models can often be described by distributions of local effective fields, whereas the unbounded continuous variables and general nonlinear convex losses considered here make the local effective problems themselves random functions. Moreover, heavy-tailed features produce a few strong contributions together with infinitely many small ones, making the order parameters probability distributions on function spaces. We derive the corresponding saddle-point equations, and the same approach may extend to other general heterogeneous continuous-variable systems. Further details are provided in Appendix A.

Learning from heavy-tailed data. Learning from entry-wise i.i.d. heavy-tailed data with squared-loss ridge regression reduces to a resolvent calculation for a heavy-tailed sample covariance matrix. The spectral distributions and resolvents of such matrices have been studied (Belinschi et al., 2009; Jung, 2018). In this quadratic setting, the local effective problems can be represented by finitely many coefficients, whereas general nonlinear convex losses require an order parameter that specifies the distribution of the entire random function. Our analysis extends this quadratic description to general nonlinear convex losses and characterizes the limiting estimator distribution and prediction error.

3 NOTATION AND PROBLEM SETUP

We study high-dimensional linear regression with coordinate-wise independent symmetric α-stable features using regularized ERM.

3.1 NOTATION

We denote by $S\alpha S(\sigma)$ the symmetric α-stable distribution with scale parameter σ, whose characteristic function is $\mathbb{E}\exp(itX) = \exp(-\sigma^\alpha|t|^\alpha)$ where $X \sim S\alpha S(\sigma)$. For a random vector $\boldsymbol{x} = (x_1, \ldots, x_D)^\top$, we write $\boldsymbol{x} \sim P^{\otimes D}$ when its coordinates are independently distributed according to P. We use $\xrightarrow{\mathrm{p}}$ and $\xrightarrow{\mathrm{d}}$ to denote convergence in probability and convergence in distribution, respectively. At a nondifferentiable point, $f'(x)$ denotes a subgradient of f at x, and $\partial f(x)$ denotes the subdifferential of f at x.

3.2 PROBLEM SETUP

We consider linear regression with N training examples in D dimensions. For $0 < \alpha < 2$, the training inputs are generated as

$$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \overset{\text{i.i.d.}}{\sim} (S\alpha S(1))^{\otimes D}. \tag{1}$$

Independently of the inputs, the teacher and label noise satisfy $\boldsymbol{w}^* \sim p_w^{\otimes D}$ and $\epsilon_\mu \overset{\text{i.i.d.}}{\sim} p_\epsilon$, where both distributions are centered, nondegenerate, and sub-Gaussian, and p_ϵ additionally has a bounded density. The labels are

$$y_\mu = \frac{1}{D^{1/\alpha}} \boldsymbol{x}_\mu^\top \boldsymbol{w}^* + \epsilon_\mu, \tag{2}$$

where $D^{-1/\alpha}$ keeps the linear predictor of order one. Given a convex loss ℓ, a strongly convex penalty function γ, and $\lambda > 0$, we define the regularized ERM estimator by

$$\hat{\boldsymbol{w}} := \operatorname*{argmin}_{\boldsymbol{w} \in \mathbb{R}^D} \left\{ \sum_{\mu=1}^{N} \ell\left(\frac{1}{D^{1/\alpha}} \boldsymbol{x}_\mu^\top \boldsymbol{w} - y_\mu \right) + \lambda \sum_{i=1}^{D} \gamma(w_i) \right\}. \tag{3}$$

To evaluate the learned estimator, we draw an independent test example $(\boldsymbol{x}_0, y_0)$ from the same model. Its label and prediction are

$$y_0 = \frac{1}{D^{1/\alpha}} \boldsymbol{x}_0^\top \boldsymbol{w}^* + \epsilon_0, \qquad \hat{y}_0 = \frac{1}{D^{1/\alpha}} \boldsymbol{x}_0^\top \hat{\boldsymbol{w}}. \tag{4}$$

For a test loss ℓ_{test}, we define the noise-free test prediction error and the generalization error by

$$\Delta_{N,D}^{(\alpha)} := \hat{y}_0 - (y_0 - \epsilon_0) = \frac{1}{D^{1/\alpha}} \boldsymbol{x}_0^\top (\hat{\boldsymbol{w}} - \boldsymbol{w}^*) \tag{5}$$

$$\mathcal{E}_{N,D}^{(\alpha)} := \mathbb{E}\left[\ell_{\text{test}}(\hat{y}_0 - y_0) \right], \tag{6}$$

where the expectation is over the test example. The test noise is excluded from $\Delta_{N,D}^{(\alpha)}$ but included in $\mathcal{E}_{N,D}^{(\alpha)}$. We study these quantities in the proportional high-dimensional limit:

$$N, D \to \infty, \qquad \frac{N}{D} \to \kappa \in (0, \infty). \tag{7}$$

3.3 REGULARITY CONDITIONS

We impose the following assumptions to guarantee a unique ERM minimizer and to ensure that the population loss has a well-defined minimum at zero residual.

Assumption 1 (Training loss)**.** *The training loss $\ell : \mathbb{R} \to \mathbb{R}$ is convex and is either globally Lipschitz or the squared loss $\ell(r) = r^2/2$. Assume*

$$\operatorname*{argmin}_{r \in \mathbb{R}} \mathbb{E}_\epsilon \ell(r - \epsilon) = \{0\}, \quad 0 < \frac{\mathrm{d}^2}{\mathrm{d}r^2} \mathbb{E}_\epsilon \ell(r - \epsilon) \Big|_{r=0} < \infty. \tag{8}$$

This assumption imposes the local strict convexity required for stable parameter estimation. Consequently, this assumption accommodates a wide range of standard objectives, including the squared loss, Huber and pseudo-Huber losses, and L_p losses ($1 \le p \le 2$).

4 PRECISE ASYMPTOTIC CHARACTERIZATION

Our goal is to characterize the generalization error of ERM estimator in the proportional high-dimensional limit. The key observation is that prediction on an independent test example depends on the learned coefficient errors only through a single scalar. Determining this scalar from the training problem, however, requires a function-valued order parameter.

Define the empirical α-error of the learned coefficients by

$$q_{\alpha,D} := \frac{1}{D}\sum_{i=1}^{D} |\hat{w}_i - w_i^*|^\alpha . \tag{9}$$

In the proportional high-dimensional limit, there exists a deterministic q_α such that $q_{\alpha,D} \xrightarrow{\mathrm{p}} q_\alpha$. In this limit, by the stability property of symmetric α-stable distributions, the noise-free test prediction error in Equation (5) satisfies

$$\Delta_{N,D}^{(\alpha)} \xrightarrow{\mathrm{d}} \Delta^{(\alpha)}(\kappa) := q_\alpha^{1/\alpha} S_\alpha, \tag{10}$$

where $S_\alpha \sim S\alpha S(1)$. This yields the following expression for the asymptotic generalization error.

Proposition 2 (Generalization error of ERM estimator)**.** *Let $0 < \alpha < 2$ and $S_\alpha \sim S\alpha S(1)$. Then, the asymptotic generalization error* (6) *is given by*

$$\mathcal{E}_{N,D}^{(\alpha)} \xrightarrow{\mathrm{p}} \mathcal{E}^{(\alpha)}(\kappa) := \mathbb{E}_{S_\alpha,\epsilon_0}\Big[\ell_{\mathrm{test}}\Big(q_\alpha^{1/\alpha} S_\alpha - \epsilon_0\Big)\Big]. \tag{11}$$

Proposition 2 shows that q_α is the only scalar needed to evaluate the asymptotic generalization error. To compute this scalar, we need the distribution of the coefficient errors $\hat{w}_i - w_i^*$.

For Gaussian features, the contribution of each data point to an individual coefficient is infinitesimally small, so the collective effect of the data becomes statistically uniform across all coefficients. As is well known, their combined effect, and consequently the coefficient distribution and q_α, is fully determined by a finite set of scalar equations.

However, for heavy-tailed features, a small number of exceptionally large entries retain a finite influence. Which entries are extreme and how strongly they affect the loss differ across coefficients, so their effects do not reduce to common scalar values. The coefficient-specific effective problem must therefore remain random. It consists of a random scale $\hat{Q} \in [0,\infty)$ and a standardized Gaussian fluctuation ζ, which together determine the linear *data fluctuation*, and a random convex *data response* function $\hat{\mathcal{R}} : \mathbb{R} \to \mathbb{R}$ that quantifies the nonlinear loss change induced by a perturbation of the coefficient. We denote their joint distribution by the functional order parameter $\hat{\mathcal{P}}$. Under the model in Section 3 and Assumption 1, with $0 < \alpha < 2$, draw

$$(\hat{Q}, \zeta, \hat{\mathcal{R}}) \sim \hat{\mathcal{P}}, \quad w^* \sim p_w. \tag{12}$$

Here w^* is independent of the first triple, $\hat{Q}$ is a nonnegative $\alpha/2$-stable random variable, $\zeta \sim \mathcal{N}(0,1)$ is independent of $\hat{Q}$, and $\hat{\mathcal{R}}$ is convex with $\hat{\mathcal{R}}(0) = 0$ and $0 \in \partial\hat{\mathcal{R}}(0)$. Although ζ is independent of $\hat{Q}$, it is generally dependent on $\hat{\mathcal{R}}$. Its full definition is given in Appendix B.

For each sample of the effective surrogate randomness, define $\hat{u}_i$ as the minimizer of an effective objective consisting of the change in regularization, a random linear fluctuation, and the nonlinear response of the training loss:

$$\hat{u}_i := \operatorname*{argmin}_{u\in\mathbb{R}} \left\{ \underbrace{\lambda[\gamma(w^*+u) - \gamma(w^*)]}_{\text{regularization}} + \underbrace{\sqrt{\hat{Q}}\,\zeta u}_{\text{data fluctuation}} + \underbrace{\hat{\mathcal{R}}(u)}_{\text{data response}} \right\}. \tag{13}$$

The $\hat{u}_i$ constructed in this way has the same limiting distribution as the learned coefficient error $\hat{w}_i - w_i^*$. The corresponding statement is the following result.

Result 3 (Statistics of the learned parameters)**.** *With $\mathbb{E}\,|\hat{u}_i|^\alpha < \infty$, the empirical coefficient distribution and its α-moment satisfy*

$$\hat{w}_i - w_i^* \xrightarrow{\mathrm{d}} \hat{u}_i \tag{14}$$

$$q_{\alpha,D} \xrightarrow{\mathrm{p}} q_\alpha := \mathbb{E}\,|\hat{u}_i|^\alpha , \tag{15}$$

where the expectation is over the joint distribution of ζ, $\hat{\mathcal{R}}$, $\hat{Q}$, and w^*.

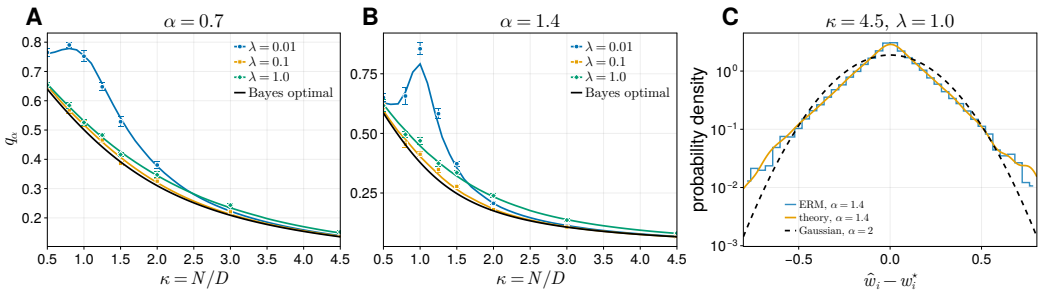

Figure 1: **Empirical validation of the theoretical coefficient-error predictions.** (A, B) Solid colored lines show the theoretical predictions of q_α (Equation 15), while markers represent numerical results obtained via finite-dimensional ERM. The solid black line shows the theoretical prediction of q_α under the Bayes–optimal setting in Section 5.3. (C) Comparison between the theoretical distribution (Equation 14) and the empirical coefficient-error distribution obtained by finite-dimensional ERM. The dashed black curve shows the theoretical coefficient-error distribution at $\alpha = 2$, which is Gaussian. (A-C) Huber loss with $\delta = 1$, $p_w = \mathcal{N}(0, 1)$, and $p_\epsilon = \mathcal{N}(0, 0.5^2)$; (A,B) $D = 256$; (C) $D = 4096$, $\lambda = 1$, and the empirical distribution is averaged over 8 independent trials. (A,B) Error bars indicate the standard error over 8 independent trials for each point.

The detailed derivation is provided in Appendix C. Figure 1 confirms that both the predicted α-moment and the full coefficient-error distribution closely match finite-dimensional ERM simulations. Thus, $\hat{\mathcal{P}}$ determines both the full limiting coefficient-error distribution and its scalar summary q_α. Combined with Equation (11), this result provides a precise characterization of the generalization error on test data.

The decomposition in Equation (13) further clarifies how heavy-tailed features affect each learned coefficient. The regularization term penalizes coefficient values according to γ, while the linear data-fluctuation term generates a random gradient that can cause the coefficient $\hat{w}_i$ to deviate from its true value w_i^*, thereby increasing the estimation error $|\hat{w}_i - w_i^*|$. The nonlinear data-response term counteracts this deviation and tends to keep the coefficient error near zero. Because the latter two terms are generated from the same heavy-tailed samples, they are statistically dependent, and their magnitudes vary across coordinates. Consequently, an extreme data fluctuation can produce a large linear term. If $\hat{\mathcal{R}}(u)$ increases sufficiently away from $u = 0$, the minimizer $\hat{u}_i$ remains close to zero; otherwise, the coefficient error can be large and increase the generalization error.

Comparison with Gaussian features. When $\alpha = 2$, the Gaussian counterpart of the heavy-tailed coefficient problem in Equation (13) is

$$\hat{u}_{\mathrm{G}} := \underset{u \in \mathbb{R}}{\operatorname{argmin}} \left\{ \underbrace{\lambda[\gamma(w^* + u) - \gamma(w^*)]}_{\text{regularization}} + \underbrace{\sqrt{\hat{q}}\,\xi u}_{\text{data fluctuation}} + \underbrace{\frac{\hat{\chi}}{2} u^2}_{\text{data response}} \right\}, \tag{16}$$

where $\xi \sim \mathcal{N}(0, 1)$ and $\hat{q}$ and $\hat{\chi}$ are some positive constants. Gaussian features average these effects into a common fluctuation scale $\sqrt{\hat{q}}$ and a common quadratic data response $\hat{\chi}u^2/2$. For heavy-tailed features, by contrast, the relative magnitudes of the linear and nonlinear terms vary randomly across coordinates, so some coefficients have small errors whereas others are poorly estimated within the same model.

5 Broader Implications of the Asymptotic Theory

Result 3 in Section 4 characterized the asymptotic coefficient distribution and generalization error of regularized ERM with entry-wise i.i.d. α-stable features. Using this characterization, we next investigate three questions that have been extensively studied for Gaussian features: universality, the scaling of the typical and extreme error, and the performance attainable by Bayes-optimal inference.

5.1 UNIVERSALITY LAW

A central question in high-dimensional learning is which properties of the feature distribution determine the asymptotic performance of trained models. For finite-variance features, Gaussian universality states that, under suitable regularity conditions, replacing the feature distribution with any other distribution matching the same mean and covariance leaves the asymptotic generalization error unchanged. The following proposition formalizes this classical result.

Proposition 4 (Gaussian universality for regularized ERM)**.** *Let the feature coordinates be i.i.d. from a sub-Gaussian distribution P_X satisfying $\mathbb{E}_{P_X} X = 0$ and $\mathbb{E}_{P_X} X^2 = 1$, and suppose that Assumption 1 holds with $\alpha = 2$. Let $q_2^{P_X}$ and q_2^{G} denote the limiting coefficient-error moments under P_X and standard Gaussian features, respectively. Then*

$$q_2^{P_X} = q_2^{\mathrm{G}}. \tag{17}$$

Thus, within the sub-Gaussian class, the limiting coefficient error, and therefore the generalization error, depends on the feature distribution only through its mean and variance. While Proposition 4 was established rigorously by Montanari & Saeed (2022), an equivalent moment universality was already classically known via non-rigorous replica calculations for finite-variance patterns (Seung et al., 1992; Watkin et al., 1993).

However, when the feature variance is infinite, this Gaussian universality law no longer applies. Nevertheless, we claim that an analogous universality principle holds through replica calculations for coordinate-wise i.i.d. heavy-tailed feature distributions.

Result 5 (Heavy-tail universality for regularized ERM)**.** *Let $0 < \alpha < 2$, and let the feature coordinates be i.i.d. from a symmetric distribution P_X satisfying*

$$\Pr(|X| > x) = C_\alpha^{\mathrm{S}} x^{-\alpha} + o\big(x^{-\alpha}\big) \qquad (x \to \infty), \tag{18}$$

where $C_\alpha^{\mathrm{S}} := 2\Gamma(\alpha)\sin(\pi\alpha/2)/\pi$ is the total tail constant of $S\alpha S(1)$ [1], and suppose that Assumption 1 holds. Let $q_\alpha^{P_X}$ and q_α^{S} denote the limiting coefficient-error moments under P_X and $S\alpha S(1)$ features, respectively. Then

$$q_\alpha^{P_X} = q_\alpha^{\mathrm{S}}. \tag{19}$$

The detailed derivation is provided in Appendix D. Result 5 demonstrates that, for the broad class of losses under Assumption 1, the limiting coefficient error and generalization error depend on the feature distribution solely through its tail index α and total tail constant. This universality principle is numerically confirmed in Figure 2.

Together, Proposition 4 and Result 5 establish two distinct universality classes: the finite-variance regime, represented by the Gaussian distribution and determined by the first two moments; and the infinite-variance regime, represented by the α-stable distribution and specified by the tail index and total tail constant. In both cases, these parameters completely determine the dependence of the limiting coefficient error on the feature distribution.

5.2 SCALING LAWS AND PREDICTION RELIABILITY

How does prediction performance improve as the training sample size increases? In this subsection, we evaluate this improvement as the sampling ratio $\kappa = N/D$ grows, using two key metrics: the typical test prediction error and the probability of an error exceeding a fixed threshold. According to Proposition 2, the prediction error induced by the estimator $\hat{\boldsymbol{w}}$ converges in distribution to $q_\alpha(\kappa)^{1/\alpha} S_\alpha$. Using this convergence, we derive the asymptotic expansion of $q_\alpha(\kappa)$ in the regime $\kappa \gg 1$ and obtain the following main results on typical performance and tail error probabilities in comparison with the finite-variance (Gaussian) baseline.

Result 6 (Typical prediction error and tail reliability with many training examples)**.** *Suppose that Assumption 1 holds. Let $\Delta^{(2)}(\kappa)$ denote the Gaussian counterpart of $\Delta^{(\alpha)}(\kappa)$ in Equation* (5)*. Then, for every $0 < \alpha < 2$, there exists a constant $c_\alpha \in (0, \infty)$ such that*

$$q_\alpha(\kappa) = \frac{c_\alpha}{\kappa} + o(\kappa^{-1}), \qquad \kappa \to \infty. \tag{20}$$

[1] If the original distribution has a total tail constant $C > 0$, rescaling X by $(C_\alpha^{\mathrm{S}}/C)^{1/\alpha}$ recovers this normalization; equivalently, the original distribution is compared against $S\alpha S((C/C_\alpha^{\mathrm{S}})^{1/\alpha})$.

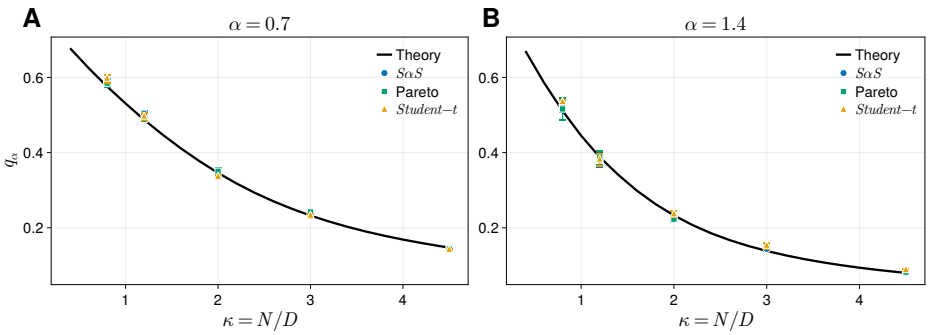

Figure 2: **Empirical validation of universality across heavy-tailed feature distributions.** The solid line shows the theoretical prediction of q_α for $S\alpha S(1)$ features, while markers represent numerical results obtained via Huber–ridge ERM with symmetric α-stable, Pareto, and Student-t coordinates. The Pareto and Student-t distributions are rescaled to satisfy $\Pr(|X| > x) \sim C_\alpha^{\mathrm{S}} x^{-\alpha}$. Parameters: (A, B) $D = 256$, $\delta = 1$, $\lambda = 1$, $p_w = \mathcal{N}(0, 1)$, and $p_\epsilon = \mathcal{N}(0, 0.5^2)$. Error bars indicate the standard error over 6 independent trials for each point.

Consequently, for every fixed $t > 0$,

$$\operatorname{median}\Big(\Big|\Delta^{(\alpha)}(\kappa)\Big|\Big) = \Theta\Big(\kappa^{-1/\alpha}\Big), \quad \Pr\Big(\Big|\Delta^{(\alpha)}(\kappa)\Big| > t\Big) = \Theta\big(\kappa^{-1}\big). \tag{21}$$

For Gaussian features,

$$\operatorname{median}\Big(\Big|\Delta^{(2)}(\kappa)\Big|\Big) = \Theta\Big(\kappa^{-1/2}\Big), \quad \Pr\Big(\Big|\Delta^{(2)}(\kappa)\Big| > t\Big) = \exp[-\Theta(\kappa)], \tag{22}$$

where $f = \Theta(g)$ means that f and g have the same asymptotic order up to constant factors.

The detailed derivation is provided in Appendix E. Figure 3 shows that the predictions of both scaling laws agree remarkably well with finite-dimensional ERM results.

This result shows that the typical prediction error decreases faster under heavy-tailed features than under Gaussian features. Intuitively, rare training observations with large feature magnitudes contain substantial information for estimating the coefficients. In contrast, the probability of exceeding a fixed error threshold is $\Theta(\kappa^{-1})$ under heavy-tailed features, whereas it is exponentially small under Gaussian features. This is because the model can exploit the extreme coordinates observed during training, whereas an extreme value in a new test example may occur in a different coordinate and amplify the residual error in the corresponding coefficient.

5.3 Bayes-Optimal Setting

In this subsection, we study prediction under Bayes-optimal inference, where the prior and noise distributions used by the learner match the data-generating model. This setting enables a more tractable analytical characterization. Moreover, for any specified test loss, the corresponding Bayes risk lower-bounds the risk of any predictor based on the same training data, including more flexible approaches such as kernel methods, ensembles, and deep neural networks. It thereby helps distinguish limitations due to heavy-tailed features from those due to the choice of inference method.

For simplicity, we assume a Gaussian teacher prior and Gaussian label noise:

$$p_w = \mathcal{N}\big(0, \tau^2\big), \quad p_\epsilon = \mathcal{N}\big(0, \sigma^2\big). \tag{23}$$

Combining the Gaussian likelihood with this prior gives

$$p(\boldsymbol{w} \mid X, \boldsymbol{y}) \propto \exp\left\{-\frac{1}{2\sigma^2}\left\|\boldsymbol{y} - D^{-1/\alpha} X\boldsymbol{w}\right\|_2^2 - \frac{1}{2\tau^2}\|\boldsymbol{w}\|_2^2\right\}, \tag{24}$$

where $X \in \mathbb{R}^{N \times D}$ has rows $\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_N^\top$, and $\boldsymbol{y} = (y_1, \ldots, y_N)^\top$ is the label vector. Completing the square shows that the posterior is Gaussian:

$$\boldsymbol{w}^* \mid X, \boldsymbol{y} \sim \mathcal{N}(\hat{\boldsymbol{w}}_{\mathrm{B}}, \Sigma_{\mathrm{B}}), \tag{25}$$

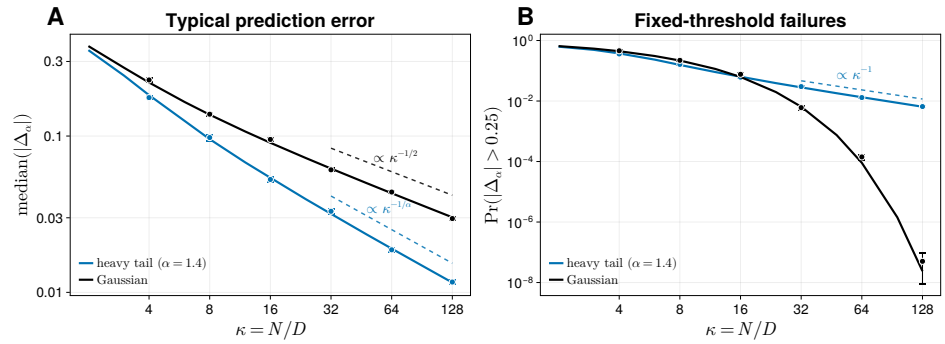

Figure 3: **Empirical validation of prediction-error scaling and reliability.** (A) The solid colored and black lines show the theoretical predictions for $\mathrm{median}(|\Delta^{(\alpha)}(\kappa)|)$ and $\mathrm{median}(|\Delta^{(2)}(\kappa)|)$ under heavy-tailed and Gaussian features, respectively, while markers represent the finite-dimensional ERM results. (B) The solid colored and black lines show the theoretical predictions for $\Pr(|\Delta^{(\alpha)}(\kappa)| > t)$ and $\Pr(|\Delta^{(2)}(\kappa)| > t)$ under heavy-tailed and Gaussian features, respectively, while markers represent the finite-dimensional ERM results. Parameters: (A, B) Huber loss with $\delta = 1$, $\lambda = 1$, $p_w = \mathcal{N}(0, 1)$, $p_\epsilon = \mathcal{N}(0, 0.5^2)$, and $D = 128$. Error bars indicate the standard error over 8 independent trials for each point.

where the posterior mean and covariance are

$$\hat{\boldsymbol{w}}_{\mathrm{B}} := \mathbb{E}[\boldsymbol{w}^* \mid X, \boldsymbol{y}] = \left(D^{-2/\alpha} X^\top X + \frac{\sigma^2}{\tau^2} I_D\right)^{-1} D^{-1/\alpha} X^\top \boldsymbol{y} \tag{26}$$

$$\Sigma_{\mathrm{B}} := \sigma^2 \left(D^{-2/\alpha} X^\top X + \frac{\sigma^2}{\tau^2} I_D\right)^{-1}. \tag{27}$$

Thus $\hat{\boldsymbol{w}}_{\mathrm{B}}$ is the squared-loss ridge estimator with $\lambda_{\mathrm{B}} = \sigma^2/\tau^2$. We introduce its proportional-limit coefficient error as

$$q_{\alpha,\mathrm{B}}(\kappa) := \mathbb{E}\left[\frac{1}{D}\sum_{i=1}^{D} |\hat{w}_{\mathrm{B},i} - w_i^*|^\alpha\right]. \tag{28}$$

When $\alpha = 2$, the coefficient error is determined by the average posterior variance and can be analyzed through the usual Stieltjes transform of the empirical covariance matrix. For $0 < \alpha < 2$, however, the error also depends on how the posterior variances vary across coordinates. Hence the normalized trace alone is insufficient, and we must characterize the distribution of the local diagonal resolvent entries. The required local-resolvent analysis nevertheless reduces to a scalar fixed-point equation for $q_{\alpha,\mathrm{B}}$. The following result states this equation together with its data-rich asymptotics.

Result 7 (Coefficient error of the Bayes ridge estimator)**.** *The coefficient error satisfies*

$$q_{\alpha,\mathrm{B}}(\kappa) = \frac{2^{\alpha/2}\Gamma((\alpha+1)/2)\sigma^\alpha}{\sqrt{\pi}\,\Gamma(1+\alpha/2)} \times \int_0^\infty \exp\left\{-\frac{\sigma^2}{\tau^2}u^{\frac{2}{\alpha}} - \frac{\kappa 2^\alpha \Gamma\left(\frac{\alpha+1}{2}\right)}{\sqrt{\pi}\,\Gamma(1+\alpha/2)} u \int_0^\infty \exp\left(-v^{\frac{2}{\alpha}} - \frac{2^{\frac{\alpha}{2}} q_{\alpha,\mathrm{B}}(\kappa)}{\sigma^\alpha} v\right) \mathrm{d}v\right\} \mathrm{d}u. \tag{29}$$

In particular, expanding Equation (29) *for large* κ *gives*

$$q_{\alpha,\mathrm{B}}(\kappa) = \frac{2^{-\alpha/2}}{\Gamma(1+\alpha/2)} \frac{\sigma^\alpha}{\kappa} + o\big(\kappa^{-1}\big). \tag{30}$$

The detailed derivation is provided in Appendix F. The black curves in Figures 1A and B show the theoretical Bayes-optimal error, which indeed provides a tight lower bound for the errors achieved by the Huber-loss estimator.

Comparison with elliptical heavy-tailed features. Following Adomaityte et al. (2024), we consider elliptical features written in our scaling convention as

$$\boldsymbol{x}_\mu^{\mathrm{ell}} = D^{1/\alpha} R_\mu \boldsymbol{z}_\mu, \qquad \boldsymbol{z}_\mu \sim \mathcal{N}(\mathbf{0}, I_D/D), \tag{31}$$

The labels are generated from Equation (2), with $\boldsymbol{x}_\mu$ replaced by $\boldsymbol{x}_\mu^{\mathrm{ell}}$. To isolate the effect of feature dependence, we choose R_μ^2, independently of $\boldsymbol{z}_\mu$, to have the limiting distribution of the normalized squared row norm in the coordinate-wise independent model. Then

$$D^{-2/\alpha}\|\boldsymbol{x}_\mu\|_2^2 \xrightarrow[D\to\infty]{\mathrm{d}} R_\mu^2, \quad D^{-2/\alpha}\|\boldsymbol{x}_\mu^{\mathrm{ell}}\|_2^2 \xrightarrow[D\to\infty]{\mathrm{d}} R_\mu^2. \tag{32}$$

Both $\boldsymbol{x}_\mu$ and $\boldsymbol{x}_\mu^{\mathrm{ell}}$ are multivariate symmetric α-stable.

Let $q_{\alpha,\mathrm{B}}^{\mathrm{ell}}(\kappa)$ denote the corresponding coefficient α-error. Specializing the Bayes-optimal equations of Adomaityte et al. (2024) to the Gaussian prior and noise in Equation (23) gives

$$1 = \frac{1}{\tau^2}\left[\frac{\sqrt{\pi}\, q_{\alpha,\mathrm{B}}^{\mathrm{ell}}(\kappa)}{2^{\alpha/2}\Gamma((\alpha+1)/2)}\right]^{\frac{2}{\alpha}} + \kappa\left[1 - \sigma^2\int_0^\infty \exp\Big(-\sigma^2 u - 2^{\alpha/2} q_{\alpha,\mathrm{B}}^{\mathrm{ell}}(\kappa) u^{\alpha/2}\Big)\,\mathrm{d}u\right]. \tag{33}$$

Comparison with Equation (29) yields the following result.

Result 8 (Coordinate-wise versus elliptical heavy tails)**.** *For $0 < \alpha < 2$ and $\kappa > 0$,*

$$q_{\alpha,\mathrm{B}}(\kappa) < q_{\alpha,\mathrm{B}}^{\mathrm{ell}}(\kappa). \tag{34}$$

Nevertheless, their data-rich asymptotics agree including the leading coefficient:

$$q_{\alpha,\mathrm{B}}(\kappa) \sim q_{\alpha,\mathrm{B}}^{\mathrm{ell}}(\kappa) \sim \frac{2^{-\alpha/2}}{\Gamma(1+\alpha/2)}\frac{\sigma^\alpha}{\kappa}. \tag{35}$$

A derivation of Equation (33) and Result 8 is provided in Appendix F. This ordering has a simple interpretation. In the entry-wise independent model, large feature magnitudes occur at different coordinates of different samples and are therefore distributed among parameter directions. In the elliptical model, by contrast, the common radial variable R_μ scales all coordinates of a sample simultaneously. At finite κ, large feature magnitudes are therefore concentrated in fewer samples and parameter directions, leading to a larger coefficient error in the elliptical model. As $\kappa \to \infty$, the number of independently observed directions is sufficient for this difference not to affect the leading-order error, so the two models have the same $1/\kappa$ asymptotics.

6 Discussion

We developed an asymptotic theory for regularized ERM in linear regression with independent symmetric α-stable feature coordinates. Using the replica method with a functional order parameter, we characterized the coefficient-error distribution and generalization error in the proportional high-dimensional limit. This characterization reveals a heavy-tail universality law: the asymptotic prediction behavior is determined only by the tail index and scale of the feature distribution. The typical prediction errors exhibit the favorable scaling $\kappa^{-1/\alpha}$, where κ denotes the number of data per dimension, whereas the probability of a fixed-magnitude error decays only as κ^{-1}. The theory also characterizes Bayes-optimal performance and clarifies how extreme observations distinguish entry-wise and elliptical heavy-tailed models.

A promising direction for future work is to develop the present results into inference algorithms for heavy-tailed data. In particular, it would be interesting to derive an approximate message passing algorithm (Kabashima, 2003; Donoho et al., 2009). Such an algorithm would provide a computationally efficient iterative inference procedure, whose macroscopic properties at fixed-point are expected to be described by the replica saddle-point equations derived in this work. More broadly, the functional order parameter provides a framework for high-dimensional systems in which local environments remain random. Retaining the distribution of these environments may enable precise asymptotic analyses of other spin glass, learning, and optimization problems with persistent local heterogeneity.

AI USE STATEMENT

Generative AI tools assisted with manuscript drafting, editing, translation, paper organization, numerical code, and the exploration and checking of mathematical derivations. All material was reviewed and verified by the authors, who take responsibility for the final manuscript.

REPRODUCIBILITY STATEMENT

Detailed derivations are provided in the appendices, and code for reproducing the numerical experiments is included in the supplementary material.

REFERENCES

Urte Adomaityte, Gabriele Sicuro, and Pierpaolo Vivo. Classification of heavy-tailed features in high dimensions: a superstatistical approach. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 43880–43893. Curran Associates, Inc., 2023. doi: 10.52202/075280-1903. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/88be023075a5a3ff3dc3b5d26623fa22-Paper-Conference.pdf.

Urte Adomaityte, Leonardo Defilippis, Bruno Loureiro, and Gabriele Sicuro. High-dimensional robust regression under heavy-tailed data: asymptotics and universality. *Journal of Statistical Mechanics: Theory and Experiment*, 2024(11):114002, nov 2024. doi: 10.1088/1742-5468/ad65e6. URL https://doi.org/10.1088/1742-5468/ad65e6.

Serban Belinschi, Amir Dembo, and Alice Guionnet. Spectral measure of heavy tailed band and covariance random matrices. *Communications in Mathematical Physics*, 289(3):1023–1055, 2009. doi: 10.1007/s00220-009-0822-4.

Christian Brownlees, Emilien Joly, and Gábor Lugosi. Empirical risk minimization for heavy-tailed losses. *The Annals of Statistics*, 43(6):2507–2536, December 2015. ISSN 0090-5364. doi: 10.1214/15-AOS1350. URL https://doi.org/10.1214/15-AOS1350.

Patrick Charbonneau, Enzo Marinari, Marc Mézard, Giorgio Parisi, Federico Ricci-Tersenghi, Gabriele Sicuro, and Francesco Zamponi (eds.). *Spin Glass Theory and Far Beyond: Replica Symmetry Breaking After 40 Years*. World Scientific, 2023. doi: 10.1142/13341.

Aditya Chopra and Brian L. Evans. Joint statistics of radio frequency interference in multi-antenna receivers. *IEEE Transactions on Signal Processing*, 60(7):3588–3603, 2012. doi: 10.1109/TSP.2012.2192431.

Luca Maria Del Bono, Flavio Nicoletti, and Federico Ricci-Tersenghi. Analytical solution to Heisenberg spin glass models on sparse random graphs and their de Almeida–Thouless line. *Physical Review B*, 110(18):184205, 2024. doi: 10.1103/PhysRevB.110.184205.

David Donoho and Andrea Montanari. High dimensional robust M-estimation: Asymptotic variance via approximate message passing. *Probability Theory and Related Fields*, 166(3–4):935–969, 2016. doi: 10.1007/s00440-015-0675-z.

David L. Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009. doi: 10.1073/pnas.0909892106. URL https://doi.org/10.1073/pnas.0909892106.

Noureddine El Karoui, Derek Bean, Peter J. Bickel, Chinghway Lim, and Bin Yu. On robust regression with high-dimensional predictors. *Proceedings of the National Academy of Sciences*, 110(36):14557–14562, 2013. doi: 10.1073/pnas.1307842110. URL https://www.pnas.org/doi/abs/10.1073/pnas.1307842110.

Silvio Franz, Michele Leone, Federico Ricci-Tersenghi, and Riccardo Zecchina. Exact solutions for diluted spin glasses and optimization problems. *Physical Review Letters*, 87(12):127209, August 2001. ISSN 1079-7114. doi: 10.1103/PhysRevLett.87.127209. URL https://doi.org/10.1103/PhysRevLett.87.127209.

Chikara Furusawa and Kunihiko Kaneko. Zipf's law in gene expression. *Phys. Rev. Lett.*, 90:088102, Feb 2003. doi: 10.1103/PhysRevLett.90.088102. URL https://link.aps.org/doi/10.1103/PhysRevLett.90.088102.

J. Ilow and D. Hatzinakos. Analytic alpha-stable noise modeling in a Poisson field of interferers or scatterers. *IEEE Transactions on Signal Processing*, 46(6):1601–1611, 1998. doi: 10.1109/78.678475.

Paul Jung. Lévy–Khintchine random matrices and the Poisson weighted infinite skeleton tree. *Transactions of the American Mathematical Society*, 370(1):641–668, 2018. doi: 10.1090/tran/6977.

Yoshiyuki Kabashima. A cdma multiuser detection algorithm on the basis of belief propagation. *Journal of Physics A: Mathematical and General*, 36(43):11111, oct 2003. doi: 10.1088/0305-4470/36/43/030. URL https://doi.org/10.1088/0305-4470/36/43/030.

Oktay Karakuş, Ercan E. Kuruoğlu, and Mustafa A. Altınkaya. Modelling impulsive noise in indoor powerline communication systems. *Signal, Image and Video Processing*, 14(8):1655–1661, 2020. doi: 10.1007/s11760-020-01708-1.

Guillaume Lecué and Matthieu Lerasle. Robust machine learning by median-of-means: theory and practice. *The Annals of Statistics*, 48(2):906–931, 2020. doi: 10.1214/19-AOS1828. URL https://doi.org/10.1214/19-AOS1828.

Guillaume Lecué and Shahar Mendelson. Regularization and the small-ball method I: sparse recovery. *The Annals of Statistics*, 46(2):611–641, 2018. doi: 10.1214/17-AOS1562. URL https://doi.org/10.1214/17-AOS1562.

Cosimo Lupo and Federico Ricci-Tersenghi. Approximating the XY model on a random graph with a q-state clock model. *Physical Review B*, 95(5):054433, 2017. doi: 10.1103/PhysRevB.95.054433.

Benoit Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419, 1963. ISSN 00219398, 15375374. URL http://www.jstor.org/stable/2350970.

Shahar Mendelson. Learning without concentration. In Maria Florina Balcan, Vitaly Feldman, and Csaba Szepesvári (eds.), *Proceedings of The 27th Conference on Learning Theory*, volume 35 of *Proceedings of Machine Learning Research*, pp. 25–39, Barcelona, Spain, 13–15 Jun 2014. PMLR. URL https://proceedings.mlr.press/v35/mendelson14.html.

Shahar Mendelson. Learning without concentration for general loss functions. *Probability Theory and Related Fields*, 171(1–2):459–502, 2018. doi: 10.1007/s00440-017-0784-y.

Marc Mézard and Giorgio Parisi. The Bethe lattice spin glass revisited. *The European Physical Journal B*, 20(2):217–233, 2001. doi: 10.1007/PL00011099.

Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications*, volume 9 of *World Scientific Lecture Notes in Physics*. World Scientific, 1987. doi: 10.1142/0271. URL https://doi.org/10.1142/0271.

David Middleton. Statistical-physical models of electromagnetic interference. *IEEE Transactions on Electromagnetic Compatibility*, EMC-19(3):106–127, 1977. doi: 10.1109/TEMC.1977.303527.

Andrea Montanari and Basil N. Saeed. Universality of empirical risk minimization. In Po-Ling Loh and Maxim Raginsky (eds.), *Proceedings of the 35th Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pp. 4310–4312. PMLR, 2022. URL https://proceedings.mlr.press/v178/montanari22a.html. Full version: arXiv:2202.08832.

I. Neri, F. L. Metz, and D. Bollé. The phase diagram of Lévy spin glasses. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(1):P01010, 2010. doi: 10.1088/1742-5468/2010/01/P01010.

Tim Rogers, Isaac Pérez Castillo, Reimer Kühn, and Koujin Takeda. Cavity approach to the spectral density of sparse symmetric random matrices. *Physical Review E*, 78(3):031116, 2008. doi: 10.1103/PhysRevE.78.031116.

H. S. Seung, H. Sompolinsky, and N. Tishby. Statistical mechanics of learning from examples. *Phys. Rev. A*, 45:6056–6091, Apr 1992. doi: 10.1103/PhysRevA.45.6056. URL https://link.aps.org/doi/10.1103/PhysRevA.45.6056.

Z.-N. Shen and G. Qin. A study of cosmic ray flux based on the noise in raw CCD data from solar images. *Journal of Geophysical Research: Space Physics*, 121(11):10712–10727, 2016. doi: 10.1002/2016JA023376. URL https://doi.org/10.1002/2016JA023376.

H. Sompolinsky, N. Tishby, and H. S. Seung. Learning from examples in large neural networks. *Phys. Rev. Lett.*, 65:1683–1686, Sep 1990. doi: 10.1103/PhysRevLett.65.1683. URL https://link.aps.org/doi/10.1103/PhysRevLett.65.1683.

Christos Thrampoulidis, Ehsan Abbasi, and Babak Hassibi. Precise error analysis of regularized M-estimators in high dimensions. *IEEE Transactions on Information Theory*, 64(8):5592–5628, 2018. doi: 10.1109/TIT.2018.2840720.

Martin J. Wainwright and Eero P. Simoncelli. Scale mixtures of Gaussians and the statistics of natural images. In S. Solla, T. Leen, and K. Müller (eds.), *Advances in Neural Information Processing Systems*, volume 12, pp. 855–861. MIT Press, 1999. URL https://proceedings.neurips.cc/paper_files/paper/1999/file/6a5dfac4be1502501489fc0f5a24b667-Paper.pdf.

Timothy L. H. Watkin, Albrecht Rau, and Michael Biehl. The statistical mechanics of learning a rule. *Rev. Mod. Phys.*, 65:499–556, Apr 1993. doi: 10.1103/RevModPhys.65.499. URL https://link.aps.org/doi/10.1103/RevModPhys.65.499.

K. Y. Michael Wong and David Saad. Equilibration through local information exchange in networks. *Physical Review E*, 74(1):010104, 2006. doi: 10.1103/PhysRevE.74.010104.

K. Y. Michael Wong and David Saad. Inference and optimization of real edges on sparse graphs: A statistical physics perspective. *Physical Review E*, 76(1):011115, 2007. doi: 10.1103/PhysRevE.76.011115.

K. Y. Michael Wong, David Saad, and Zhuo Gao. Message passing for task redistribution on sparse graphs. In Y. Weiss, B. Schölkopf, and J. Platt (eds.), *Advances in Neural Information Processing Systems*, volume 18, pp. 1529–1536. MIT Press, 2005. URL https://proceedings.neurips.cc/paper_files/paper/2005/file/dc20d1211f3e7a99d775b26052e0163e-Paper.pdf.

A Additional Related Work

The functional order parameters used in this work were first developed for heterogeneous systems whose coordinate-specific effective quantities remain nonuniform even in the high-dimensional limit. In spin glasses, optimization problems, and random matrices with sparse or heavy-tailed interactions, this heterogeneity cannot generally be summarized by finitely many scalar order parameters; instead, the distribution of the coordinate-specific effective quantities must be specified (Mézard & Parisi, 2001; Franz et al., 2001; Neri et al., 2010; Jung, 2018).

In discrete-variable models such as diluted spin glasses, local effects can often be parameterized by finitely many effective fields, so that the order parameter is a probability distribution over these fields (Mézard & Parisi, 2001; Franz et al., 2001). For continuous-variable models, when the local effective problem cannot be represented by a finite-dimensional family, it must instead be represented as a function of the continuous variable, and the order parameter becomes a probability distribution on a function space. Previous studies have primarily considered cases in which the effective functions can be represented by finitely many coefficients because of a quadratic structure, their domains are compact and hence amenable to numerical discretization, or each variable has only finitely many

interactions in a strictly sparse model (Rogers et al., 2008; Lupo & Ricci-Tersenghi, 2017; Del Bono et al., 2024; Wong et al., 2005; Wong & Saad, 2006; 2007).

In contrast, our model combines general convex losses and unbounded continuous variables with heavy-tailed interactions, for which there are a few contributions of large magnitude and infinitely many of small magnitude. We derive probability distributions on function spaces and their saddle-point equations, whose solutions characterize the limiting coefficient distribution and prediction error. This construction provides a basis for analyzing continuous-variable systems with infinitely many heterogeneous interactions.

B Self-Consistent Equations

B.1 Gaussian reference equations

For comparison, we state only the resulting scalar saddle-point equations for Gaussian features, with the feature variance included in the definitions of the order parameters. Let $z, \hat{z} \sim \mathcal{N}(0,1)$ be independent of each other and of $w^* \sim p_w$ and $\epsilon \sim p_\epsilon$. Introduce an external field h and define

$$u_*(h) := \operatorname*{argmin}_{u\in\mathbb{R}} \left\{ \lambda[\gamma(w^*+u) - \gamma(w^*)] + \frac{\hat{\chi}}{2}u^2 + \sqrt{\hat{q}}\,\hat{z}u - hu \right\}. \tag{36}$$

The coefficient-side saddle equations are

$$q = \mathbb{E}\Big[u_*(0)^2\Big] \tag{37}$$

$$\chi = \mathbb{E}\left[\frac{\partial u_*(h)}{\partial h}\bigg|_{h=0} \right]. \tag{38}$$

Using h again as a local perturbation variable, the sample-side problem is

$$v_*(h) := \operatorname*{argmin}_{v\in\mathbb{R}} \left\{ \ell(v) + \frac{\big(v - \sqrt{q}\,z + \epsilon - h\big)^2}{2\chi} \right\}. \tag{39}$$

Its saddle equations are

$$\hat{q} = \kappa\mathbb{E}\Big[\ell'(v_*(0))^2\Big] \tag{40}$$

$$\hat{\chi} = \kappa\mathbb{E}\left[\frac{\partial}{\partial h}\ell'(v_*(h))\bigg|_{h=0} \right]. \tag{41}$$

Thus the two scalar problems define the coupled updates

$$(\hat{q},\hat{\chi}) \xrightarrow{u_*(h)} (q,\chi) \tag{42}$$

$$(q,\chi) \xrightarrow{v_*(h)} (\hat{q},\hat{\chi}). \tag{43}$$

The heavy-tailed equations below also consist of two problems, but use random nonlinear response functions instead of scalar quadratic responses.

B.2 Poisson point processes

Define the convex conjugate of ℓ by

$$\ell^*(p) := \sup_{r\in\mathbb{R}}\{pr - \ell(r)\}, \tag{44}$$

and the effective domain of ℓ^* by

$$K_\ell := \{p \in \mathbb{R} : \ell^*(p) < \infty\} = \operatorname{dom}\ell^*. \tag{45}$$

Set

$$d_\alpha := \frac{\Gamma(\alpha+1)\sin(\pi\alpha/2)}{2^{\alpha/2}\sqrt{\pi}\,\Gamma((\alpha+1)/2)}. \tag{46}$$

We use two independent Poisson point processes $\{s_j\}_{j\geq 1}$ and $\{\hat{s}_l\}_{l\geq 1}$ on $(0,\infty)$. Their points, ordered by decreasing magnitude, satisfy

$$\#\{j : s_j > t\} \overset{\mathrm{d}}{=} \mathrm{Poisson}\left(\frac{2d_\alpha}{\alpha} t^{-\alpha/2}\right), \tag{47}$$

$$\#\{l : \hat{s}_l > t\} \overset{\mathrm{d}}{=} \mathrm{Poisson}\left(\frac{2\kappa d_\alpha}{\alpha} t^{-\alpha/2}\right) \tag{48}$$

for $t > 0$. Independently, let

$$g_j, \hat{g}_l \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1). \tag{49}$$

These are independent Gaussian variables attached to the Poisson points s_j and $\hat{s}_l$, respectively; such attached variables are called *marks*.

B.3 Fixed-point equation

Fix $0 < \alpha < 2$ and suppose that Assumption 1 holds. We seek a pair of probability distributions $\mathcal{P}$ for $(Q, \xi, \mathcal{R})$ and $\hat{\mathcal{P}}$ for $(\hat{Q}, \zeta, \hat{\mathcal{R}})$, such that each distribution is generated from the other by the following two updates.

B.3.1 From $\hat{\mathcal{P}}$ to $\mathcal{P}$

Here $\hat{Q} \geq 0$, and $\hat{\mathcal{R}}$ is a random convex function satisfying $\hat{\mathcal{R}}(0) = 0$ and $0 \in \partial\hat{\mathcal{R}}(0)$. Draw independent copies

$$\left(\hat{Q}_j, \zeta_j, \hat{\mathcal{R}}_j\right) \overset{\text{i.i.d.}}{\sim} \hat{\mathcal{P}} \tag{50}$$

$$w_j^* \overset{\text{i.i.d.}}{\sim} p_w \tag{51}$$

independently of $\{(s_j, g_j)\}_{j\geq 1}$, with the w_j^* independent of the triples. As in the Gaussian coefficient problem, introduce an external field h and define

$$u_{j,*}(h) := \operatorname*{argmin}_{u\in\mathbb{R}} \left\{ \lambda\big[\gamma\big(w_j^* + u\big) - \gamma\big(w_j^*\big)\big] + \sqrt{\hat{Q}_j}\,\zeta_j u + \hat{\mathcal{R}}_j(u) - hu \right\}. \tag{52}$$

The coefficient-side saddle-point equations are

$$Q = \sum_{j\geq 1} s_j u_{j,*}(0)^2 \tag{53}$$

$$\xi = \frac{1}{\sqrt{Q}} \sum_{j\geq 1} \sqrt{s_j}\, g_j u_{j,*}(0) \tag{54}$$

$$\mathcal{R}'(h) = \sum_{j\geq 1} \sqrt{s_j}\, g_j \big[u_{j,*}\big(-\sqrt{s_j}\, g_j h\big) - u_{j,*}(0)\big], \tag{55}$$

where $\mathcal{R}(0) = 0$.

The first update requires the resulting joint distribution to satisfy

$$(Q, \xi, \mathcal{R}) \sim \mathcal{P}. \tag{56}$$

B.3.2 From $\mathcal{P}$ to $\hat{\mathcal{P}}$

Draw independent copies

$$(Q_l, \xi_l, \mathcal{R}_l) \overset{\text{i.i.d.}}{\sim} \mathcal{P} \tag{57}$$

$$\epsilon_l \overset{\text{i.i.d.}}{\sim} p_\epsilon, \tag{58}$$

independently of $\{(\hat{s}_l, \hat{g}_l)\}_{l\geq 1}$. As in the Gaussian sample problem, introduce an external field h and define

$$v_{l,*}(h) \in \operatorname*{argmin}_{v\in\mathbb{R}} \left\{ \ell(v) + (-\mathcal{R}_l)^* \left(\sqrt{Q_l}\, \xi_l - \epsilon_l + h - v \right) \right\}. \tag{59}$$

The sample-side saddle-point equations are

$$\hat{Q} = \sum_{l\geq 1} \hat{s}_l \ell'(v_{l,*}(0))^2 \tag{60}$$

$$\zeta = \frac{1}{\sqrt{\hat{Q}}} \sum_{l\geq 1} \sqrt{\hat{s}_l}\, \hat{g}_l \ell'(v_{l,*}(0)) \tag{61}$$

$$\hat{\mathcal{R}}'(h) = \sum_{l\geq 1} \sqrt{\hat{s}_l}\, \hat{g}_l \Big[\ell'\Big(v_{l,*}\Big(\sqrt{\hat{s}_l}\, \hat{g}_l h\Big)\Big) - \ell'(v_{l,*}(0))\Big], \tag{62}$$

where $\hat{\mathcal{R}}(0) = 0$.

The second update requires the resulting joint distribution to satisfy

$$\Big(\hat{Q}, \zeta, \hat{\mathcal{R}}\Big) \sim \hat{\mathcal{P}}. \tag{63}$$

Equations (56) and (63) are the two coupled saddle-point equations. Thus the two random scalar problems define the coupled updates

$$\hat{\mathcal{P}} \xrightarrow{u_{j,*}(h)} \mathcal{P} \tag{64}$$

$$\mathcal{P} \xrightarrow{v_{l,*}(h)} \hat{\mathcal{P}}. \tag{65}$$

These equations also consist of two problems, but use the joint distributions of a random scale, a standard Gaussian field, and a random nonlinear response function instead of scalar order parameters.

B.4 Limiting coefficient and preactivation distributions

At the fixed point, draw

$$\Big(\hat{Q}_i, \zeta_i, \hat{\mathcal{R}}_i\Big) \sim \hat{\mathcal{P}} \tag{66}$$

$$w_i^* \sim p_w, \tag{67}$$

independently. The coefficient-side one-body problem is

$$u_{i,*}(h) := \operatorname*{argmin}_{u\in\mathbb{R}} \Big\{ \lambda[\gamma(w_i^* + u) - \gamma(w_i^*)] + \sqrt{\hat{Q}_i}\, \zeta_i u + \hat{\mathcal{R}}_i(u) - hu \Big\}. \tag{68}$$

Its unperturbed solution gives the limiting coefficient displacement:

$$\hat{w}_i - w_i^* \xrightarrow{\mathrm{d}} u_{i,*}(0). \tag{69}$$

For the sample side, draw

$$(Q_\mu, \xi_\mu, \mathcal{R}_\mu) \sim \mathcal{P} \tag{70}$$

$$\epsilon_\mu \sim p_\epsilon, \tag{71}$$

independently, and define

$$v_{\mu,*}(h) \in \operatorname*{argmin}_{v\in\mathbb{R}} \Big\{ \ell(v) + (-\mathcal{R}_\mu)^* \Big(\sqrt{Q_\mu}\, \xi_\mu - \epsilon_\mu + h - v\Big) \Big\}. \tag{72}$$

Its unperturbed solution gives the limiting training preactivation, or equivalently the fitted residual entering the loss:

$$\frac{1}{D^{1/\alpha}} \boldsymbol{x}_\mu^\top \hat{\boldsymbol{w}} - y_\mu = \frac{1}{D^{1/\alpha}} \boldsymbol{x}_\mu^\top (\hat{\boldsymbol{w}} - \boldsymbol{w}^*) - \epsilon_\mu \xrightarrow{\mathrm{d}} v_{\mu,*}(0). \tag{73}$$

C Replica Derivation of the Functional Fixed Point

This appendix derives the self-consistent equations in Appendix B.

C.1 REPLICATED ACTION

Set $\boldsymbol{u} = \boldsymbol{w} - \boldsymbol{w}^*$ and introduce the partition function

$$Z_\beta := \int_{\mathbb{R}^D} \mathrm{d}\boldsymbol{u} \, \exp\left\{ -\beta \sum_{\mu=1}^{N} \ell\Big(D^{-1/\alpha} \boldsymbol{x}_\mu^\top \boldsymbol{u} - \epsilon_\mu \Big) - \beta\lambda \sum_{i=1}^{D} \gamma(w_i^* + u_i) \right\}. \tag{74}$$

Its zero-temperature free energy is the optimized training objective:

$$-\lim_{\beta\to\infty} \frac{1}{\beta D} \mathbb{E} \log Z_\beta = \frac{1}{D} \mathbb{E} \min_{\boldsymbol{u}\in\mathbb{R}^D} \left\{ \sum_{\mu=1}^{N} \ell\Big(D^{-1/\alpha} \boldsymbol{x}_\mu^\top \boldsymbol{u} - \epsilon_\mu \Big) + \lambda \sum_{i=1}^{D} \gamma(w_i^* + u_i) \right\}. \tag{75}$$

The replica identity replaces the logarithm by integer moments:

$$\mathbb{E} \log Z_\beta = \lim_{n\to 0} \frac{\log \mathbb{E} Z_\beta^n}{n}. \tag{76}$$

The right-hand side is first evaluated for integer $n \geq 1$ and then analytically continued to a neighborhood of $n = 0$. From this point, our goal is to evaluate the data average of the replicated partition function Z_β^n for integer n.

Set

$$g_\beta(r) := e^{-\beta \ell(r)}. \tag{77}$$

For integer n, let u_i^a denote coordinate i ($i \in \{1, \ldots, D\}$) in replica a ($a \in \{1, \ldots, n\}$). Direct multiplication of the n copies of Equation (74) gives

$$Z_\beta^n = \int \prod_{a=1}^{n} \prod_{i=1}^{D} \mathrm{d}u_i^a \, \exp\left(-\beta\lambda \sum_{a=1}^{n} \sum_{i=1}^{D} \gamma(w_i^* + u_i^a) \right) \prod_{\mu=1}^{N} \prod_{a=1}^{n} g_\beta\left(D^{-1/\alpha} \sum_{i=1}^{D} x_{\mu i} u_i^a - \epsilon_\mu \right). \tag{78}$$

Write $\boldsymbol{u}_i = (u_i^1, \ldots, u_i^n)^\top$ and introduce the empirical measure of the replicated coordinate vectors,

$$\rho_D(\mathrm{d}\boldsymbol{u}) := \frac{1}{D} \sum_{i=1}^{D} \delta_{\boldsymbol{u}_i}(\mathrm{d}\boldsymbol{u}). \tag{79}$$

For one factor in the product over μ, let $\boldsymbol{k} = (k_1, \ldots, k_n)^\top \in \mathbb{R}^n$. Independence over the feature coordinates and the characteristic function of $S\alpha S(1)$ give

$$\mathbb{E}_{\boldsymbol{x}_\mu} \exp\left(iD^{-1/\alpha} \sum_{a=1}^{n} k_a \sum_{i=1}^{D} x_{\mu i} u_i^a \right) = \prod_{i=1}^{D} \mathbb{E}_{x_{\mu i}} \exp\left(iD^{-1/\alpha} x_{\mu i} \sum_{a=1}^{n} k_a u_i^a \right) \tag{80}$$

$$= \prod_{i=1}^{D} \exp\left(-\frac{1}{D} \left| \sum_{a=1}^{n} k_a u_i^a \right|^\alpha \right) \tag{81}$$

$$= \exp\left(-\frac{1}{D} \sum_{i=1}^{D} \left| \boldsymbol{k}^\top \boldsymbol{u}_i \right|^\alpha \right) \tag{82}$$

$$= \exp\left(-\int_{\mathbb{R}^n} \rho_D(\mathrm{d}\boldsymbol{v}) \left| \boldsymbol{k}^\top \boldsymbol{v} \right|^\alpha \right). \tag{83}$$

Use the Fourier convention

$$\widetilde{g}_\beta(k) := \int_{\mathbb{R}} \mathrm{d}r \, g_\beta(r) e^{-ikr}, \tag{84}$$

$$g_\beta(r) = \int_{\mathbb{R}} \frac{\mathrm{d}k}{2\pi} \, \widetilde{g}_\beta(k) e^{ikr}. \tag{85}$$

This step assumes that Fourier inversion is valid; otherwise the following formula is interpreted distributionally. Applying it to all n loss factors in Equation (78) and then using Equation (83) yields

$$\mathbb{E}_{\boldsymbol{x},\epsilon} \prod_{a=1}^{n} g_\beta\left(D^{-1/\alpha} \sum_{i=1}^{D} x_i u_i^a - \epsilon\right) \tag{86}$$

$$= \int_{\mathbb{R}^n} \prod_{a=1}^{n} \frac{\mathrm{d}k_a}{2\pi}\, \widetilde{g}_\beta(k_a) \mathbb{E}_\epsilon \exp\left(-i\epsilon \sum_{a=1}^{n} k_a\right) \exp\left(-\int_{\mathbb{R}^n} \rho_D(\mathrm{d}\boldsymbol{v}) \left|\boldsymbol{k}^\top \boldsymbol{v}\right|^\alpha\right) \tag{87}$$

$$=: \mathcal{I}_n[\rho_D]. \tag{88}$$

The integral in the second line, with ρ_D replaced by an arbitrary probability measure ρ on $\mathbb{R}^n$, defines $\mathcal{I}_n[\rho]$.

The training examples are independent. Therefore, conditioning on $\boldsymbol{w}^*$ and using the same empirical measure for every factor,

$$\mathbb{E}_{\{\boldsymbol{x}_\mu, \epsilon_\mu\}_{\mu=1}^N} \left[Z_\beta^n \mid \boldsymbol{w}^*\right] \tag{89}$$

$$= \int \prod_{a=1}^{n} \prod_{i=1}^{D} \mathrm{d}u_i^a \, \exp\left(-\beta\lambda \sum_{a=1}^{n} \sum_{i=1}^{D} \gamma(w_i^* + u_i^a)\right) \prod_{\mu=1}^{N} \mathcal{I}_n[\rho_D] \tag{90}$$

$$= \int \prod_{a=1}^{n} \prod_{i=1}^{D} \mathrm{d}u_i^a \, \exp\left(-\beta\lambda \sum_{a=1}^{n} \sum_{i=1}^{D} \gamma(w_i^* + u_i^a)\right) \mathcal{I}_n[\rho_D]^N. \tag{91}$$

We enforce Equation (79) before integrating over the replicated coordinates. The Fourier representation of the functional delta gives

$$1 = \int \mathcal{D}\rho\, \delta\left[D\rho - \sum_{i=1}^{D} \delta_{\boldsymbol{u}_i}\right] \propto \int \mathcal{D}\rho\, \mathcal{D}\hat{\rho}\, \exp\left\{-D \int_{\mathbb{R}^n} \rho(\mathrm{d}\boldsymbol{u}) \hat{\rho}(\boldsymbol{u}) + \sum_{i=1}^{D} \hat{\rho}(\boldsymbol{u}_i)\right\}. \tag{92}$$

The second representation initially integrates the conjugate field along an imaginary contour. The omitted proportionality factor is independent of the replicated coordinates and is subexponential in D. Define the one-coordinate integral

$$\mathcal{Z}_\gamma(w; \hat{\rho}) := \int_{\mathbb{R}^n} \mathrm{d}\boldsymbol{u} \, \exp\left\{\hat{\rho}(\boldsymbol{u}) - \beta\lambda \sum_{a=1}^{n} \gamma(w + u^a)\right\}. \tag{93}$$

Substituting Equation (92) into Equation (91) and averaging the independent teacher coordinates gives

$$\begin{aligned}
\mathbb{E} Z_\beta^n &\doteq \int \mathcal{D}\rho\, \mathcal{D}\hat{\rho}\, \exp\left\{-D \int \rho(\mathrm{d}\boldsymbol{u}) \hat{\rho}(\boldsymbol{u}) + N \log \mathcal{I}_n[\rho]\right\} \prod_{i=1}^{D} \mathbb{E}_{w_i^*} \mathcal{Z}_\gamma(w_i^*; \hat{\rho}) \\
&= \int \mathcal{D}\rho\, \mathcal{D}\hat{\rho}\, \exp\left\{-D \int \rho(\mathrm{d}\boldsymbol{u}) \hat{\rho}(\boldsymbol{u}) + N \log \mathcal{I}_n[\rho] + D \log \mathbb{E}_{w^*} \mathcal{Z}_\gamma(w^*; \hat{\rho})\right\} \\
&= \int \mathcal{D}\rho\, \mathcal{D}\hat{\rho}\, \exp(D\mathcal{S}_n[\rho, \hat{\rho}] + o(D)),
\end{aligned} \tag{94}$$

where $\doteq$ denotes equality on the exponential scale and $N/D \to \kappa$. Here we define the action as

$$\mathcal{S}_n[\rho, \hat{\rho}] \tag{95}$$

$$= -\int \rho(\mathrm{d}\boldsymbol{u}) \hat{\rho}(\boldsymbol{u}) + \kappa \log \mathcal{I}_n[\rho] + \log \mathbb{E}_{w^*} \mathcal{Z}_\gamma(w^*; \hat{\rho}) \tag{96}$$

$$= -\int \rho(\mathrm{d}\boldsymbol{u}) \hat{\rho}(\boldsymbol{u}) + \kappa \log \mathcal{I}_n[\rho] + \log \mathbb{E}_{w^*} \int_{\mathbb{R}^n} \mathrm{d}\boldsymbol{u} \, \exp\left\{\hat{\rho}(\boldsymbol{u}) - \beta\lambda \sum_{a=1}^{n} \gamma(w^* + u^a)\right\}. \tag{97}$$

Next, we take the $D \to \infty$ limit of the action. The saddle point assumption gives

$$\lim_{D\to\infty} \frac{1}{D} \log \mathbb{E} Z_\beta^n = \operatorname*{extr}_{\rho,\hat{\rho}} \mathcal{S}_n[\rho, \hat{\rho}]. \tag{98}$$

Differentiating the action with respect to $\hat{\rho}$ gives

$$\rho(\mathrm{d}\boldsymbol{u}) = \frac{\mathbb{E}_{w^*} \exp\{\hat{\rho}(\boldsymbol{u}) - \beta\lambda \sum_{a=1}^n \gamma(w^* + u^a)\}}{\mathbb{E}_{w^*} \int_{\mathbb{R}^n} \mathrm{d}\boldsymbol{v} \, \exp\{\hat{\rho}(\boldsymbol{v}) - \beta\lambda \sum_{a=1}^n \gamma(w^* + v^a)\}} \mathrm{d}\boldsymbol{u}, \tag{99}$$

and the variation with respect to ρ gives

$$\hat{\rho}(\boldsymbol{u}) = \kappa \frac{\delta \log \mathcal{I}_n[\rho]}{\delta \rho(\boldsymbol{u})} = \frac{\kappa}{\mathcal{I}_n[\rho]} \frac{\delta \mathcal{I}_n[\rho]}{\delta \rho(\boldsymbol{u})}. \tag{100}$$

For variations of ρ with zero total mass, the functional derivative is defined up to an additive constant. A representative obtained directly from Equation (88) is

$$\frac{\delta \mathcal{I}_n[\rho]}{\delta \rho(\boldsymbol{u})} = -\int_{\mathbb{R}^n} \prod_{a=1}^n \frac{\mathrm{d}k_a}{2\pi} \widetilde{g}_\beta(k_a) \mathbb{E}_\epsilon \exp\left(-i\epsilon \sum_{a=1}^n k_a\right) \left|\boldsymbol{k}^\top \boldsymbol{u}\right|^\alpha \exp\left(-\int_{\mathbb{R}^n} \rho(\mathrm{d}\boldsymbol{v}) \left|\boldsymbol{k}^\top \boldsymbol{v}\right|^\alpha\right), \tag{101}$$

and hence

$$\hat{\rho}(\boldsymbol{u}) = -\frac{\kappa}{\mathcal{I}_n[\rho]} \int_{\mathbb{R}^n} \prod_{a=1}^n \frac{\mathrm{d}k_a}{2\pi} \widetilde{g}_\beta(k_a) \mathbb{E}_\epsilon \exp\left(-i\epsilon \sum_{a=1}^n k_a\right) \left|\boldsymbol{k}^\top \boldsymbol{u}\right|^\alpha \exp\left(-\int_{\mathbb{R}^n} \rho(\mathrm{d}\boldsymbol{v}) \left|\boldsymbol{k}^\top \boldsymbol{v}\right|^\alpha\right). \tag{102}$$

Equation (99) gives $\int \rho(\mathrm{d}\boldsymbol{u}) = 1$. Consequently, for every constant c,

$$\mathcal{S}_n[\rho, \hat{\rho} + c] = \mathcal{S}_n[\rho, \hat{\rho}] - c \int \rho(\mathrm{d}\boldsymbol{u}) + c = \mathcal{S}_n[\rho, \hat{\rho}]. \tag{103}$$

Thus an additive constant in $\hat{\rho}$ is immaterial.

C.2 Replica-symmetric distributional fixed point

Equations (99) and (102) are saddle equations for functions on $\mathbb{R}^n$. To analyze these equations in the $n \to 0$ limit, we assume replica symmetry (RS). Although RS remains an assumption rather than a theorem for the present model, to our knowledge there is no known convex problem in the proportional high-dimensional limit for which it gives incorrect predictions. This expectation is also consistent with rigorous analyses of related convex estimators with Gaussian features (Thrampoulidis et al., 2018; Donoho & Montanari, 2016) and with the numerical agreement in Figure 1 for the present model. Under this assumption, the replicated coordinates are conditionally independent Gibbs draws from a common random one-dimensional objective. We write this assumption as

$$\rho(\mathrm{d}\boldsymbol{u}) := \int \Pi_\beta(\mathrm{d}\mathcal{V}) \prod_{a=1}^n \mu_{\mathcal{V}}(\mathrm{d}u^a), \tag{104}$$

$$\mu_{\mathcal{V}}(\mathrm{d}u) := \frac{e^{-\beta \mathcal{V}(u)}}{\int_{\mathbb{R}} e^{-\beta \mathcal{V}(v)} \, \mathrm{d}v} \, \mathrm{d}u. \tag{105}$$

Here Π_β is a probability distribution on random one-dimensional coefficient objectives $\mathcal{V} : \mathbb{R} \to \mathbb{R}$. Through the map

$$\mathcal{V} \longmapsto \mu_{\mathcal{V}} \in \mathcal{P}(\mathbb{R}), \tag{106}$$

it induces a probability distribution on the space $\mathcal{P}(\mathbb{R})$ of probability measures on $\mathbb{R}$. Equivalently, (104) describes the two-stage draw

$$\mathcal{V} \sim \Pi_\beta, \qquad U^1, \ldots, U^n \mid \mathcal{V} \overset{\text{i.i.d.}}{\sim} \mu_{\mathcal{V}}. \tag{107}$$

Thus all replicas share the same random Gibbs measure, and they become independent only after conditioning on that measure. Because

$$\mu_{\mathcal{V}+c} = \frac{e^{-\beta[\mathcal{V}(u)+c]}}{\int_{\mathbb{R}} e^{-\beta[\mathcal{V}(v)+c]}\,\mathrm{d}v}\,\mathrm{d}u = \frac{e^{-\beta\mathcal{V}(u)}}{\int_{\mathbb{R}} e^{-\beta\mathcal{V}(v)}\,\mathrm{d}v}\,\mathrm{d}u = \mu_{\mathcal{V}}, \tag{108}$$

the Gibbs measure determines $\mathcal{V}$ only up to an additive constant. We select a unique representative by imposing $\mathcal{V}(0) = 0$.

For the conjugate saddle field, we impose from the outset the RS ansatz on $e^{\hat{\rho}}$. Since $\hat{\rho}$ is not a probability measure, this ansatz uses unnormalized Boltzmann weights rather than normalized Gibbs measures. Namely, we assume that there is a probability law $\hat{\Pi}_\beta$ on random functions $\hat{\mathcal{V}}_\beta : \mathbb{R} \to \mathbb{R}$, normalized by $\hat{\mathcal{V}}_\beta(0) = 0$, such that

$$e^{\hat{\rho}_{\mathrm{RS}}(\boldsymbol{u})} := \int \hat{\Pi}_\beta\Big(\mathrm{d}\hat{\mathcal{V}}\Big) \prod_{a=1}^{n} e^{-\beta\hat{\mathcal{V}}(u^a)}. \tag{109}$$

Thus the two RS order parameters have the parallel generative forms

$$\mathcal{V}_\beta \sim \Pi_\beta, \qquad U^a \mid \mathcal{V}_\beta \overset{\text{i.i.d.}}{\sim} \mu_{\mathcal{V}_\beta}, \tag{110}$$

$$\hat{\mathcal{V}}_\beta \sim \hat{\Pi}_\beta, \qquad e^{\hat{\rho}_{\mathrm{RS}}(\boldsymbol{u})} = \mathbb{E}_{\hat{\mathcal{V}}_\beta} \prod_{a=1}^{n} e^{-\beta\hat{\mathcal{V}}_\beta(u^a)}. \tag{111}$$

The $\hat{\Pi}_\beta \to \Pi_\beta$ update: substitution into ρ-stationarity. We first substitute the second line of (111) into the ρ-stationarity condition (99). For independent $w^* \sim p_w$ and $\hat{\mathcal{V}}_\beta \sim \hat{\Pi}_\beta$, this gives

$$\rho(\mathrm{d}\boldsymbol{u}) = \frac{\mathbb{E}\Big[\prod_{a=1}^{n} e^{-\beta[\lambda\gamma(w^*+u^a)+\hat{\mathcal{V}}_\beta(u^a)]}\,\mathrm{d}u^a\Big]}{\mathbb{E}\Big[\prod_{a=1}^{n}\int_{\mathbb{R}} e^{-\beta[\lambda\gamma(w^*+v^a)+\hat{\mathcal{V}}_\beta(v^a)]}\,\mathrm{d}v^a\Big]}. \tag{112}$$

The exponent in (112) can be rewritten as

$$\lambda\gamma(w^*+u) + \hat{\mathcal{V}}_\beta(u) = \lambda\gamma(w^*) + \lambda[\gamma(w^*+u) - \gamma(w^*)] + \hat{\mathcal{V}}_\beta(u). \tag{113}$$

The u-dependent part on the right-hand side of (113) vanishes at $u = 0$. We identify this part with the random coefficient objective $\mathcal{V}_\beta$ already introduced in (107):

$$\mathcal{V}_\beta(u) := \lambda[\gamma(w^*+u) - \gamma(w^*)] + \hat{\mathcal{V}}_\beta(u). \tag{114}$$

In particular,

$$\mathcal{V}_\beta(0) = \lambda[\gamma(w^*) - \gamma(w^*)] + \hat{\mathcal{V}}_\beta(0) = 0. \tag{115}$$

Using (114) in (112) gives

$$\rho(\mathrm{d}\boldsymbol{u}) = \frac{\mathbb{E}\big[e^{-n\beta\lambda\gamma(w^*)}\prod_{a=1}^{n} e^{-\beta\mathcal{V}_\beta(u^a)}\,\mathrm{d}u^a\big]}{\mathbb{E}\big[e^{-n\beta\lambda\gamma(w^*)}\prod_{a=1}^{n}\int_{\mathbb{R}} e^{-\beta\mathcal{V}_\beta(v^a)}\,\mathrm{d}v^a\big]} \tag{116}$$

$$= \frac{\mathbb{E}\Big[e^{-n\beta\lambda\gamma(w^*)}\big(\int_{\mathbb{R}} e^{-\beta\mathcal{V}_\beta(v)}\,\mathrm{d}v\big)^n \prod_{a=1}^{n}\mu_{\mathcal{V}_\beta}(\mathrm{d}u^a)\Big]}{\mathbb{E}\big[e^{-n\beta\lambda\gamma(w^*)}\big(\int_{\mathbb{R}} e^{-\beta\mathcal{V}_\beta(v)}\,\mathrm{d}v\big)^n\big]} \tag{117}$$

$$\xrightarrow{n\to 0} \mathbb{E}\prod_{a=1}^{n}\mu_{\mathcal{V}_\beta}(\mathrm{d}u^a) \tag{118}$$

$$= \int \mathrm{Law}(\mathcal{V}_\beta)(\mathrm{d}\mathcal{V}) \prod_{a=1}^{n}\mu_{\mathcal{V}}(\mathrm{d}u^a). \tag{119}$$

Comparing (119) with (104) yields the first distributional relation

$$\hat{\mathcal{V}}_\beta \sim \hat{\Pi}_\beta, \quad w^* \sim p_w \quad\Longrightarrow\quad \mathcal{V}_\beta = \lambda[\gamma(w^*+\cdot) - \gamma(w^*)] + \hat{\mathcal{V}}_\beta \sim \Pi_\beta. \tag{120}$$

The $\Pi_\beta \to \hat{\Pi}_\beta$ update: substitution into $\hat{\rho}$-stationarity. We now substitute the primal RS ansatz (104) into the $\hat{\rho}$-stationarity condition (102). We first rewrite its stable exponent in a form that makes this substitution explicit. For ρ with finite α-moment, the stable exponent in Equation (102) has the representation

$$\int_{\mathbb{R}^n} \rho(\mathrm{d}\boldsymbol{v}) \left|\boldsymbol{k}^\top \boldsymbol{v}\right|^\alpha = d_\alpha \int_0^\infty \mathbb{E}_{g,\boldsymbol{v}\sim\rho}\big[1-\cos\big(\sqrt{s}\, g\, \boldsymbol{k}^\top \boldsymbol{v}\big)\big] s^{-1-\alpha/2}\,\mathrm{d}s. \tag{121}$$

Here d_α is defined in Appendix B and $g \sim \mathcal{N}(0,1)$. The identity follows from $\mathbb{E}_g \cos(\sqrt{s}\, gz) = e^{-sz^2/2}$. The following formula identifies the exponential of the negative stable exponent with the characteristic function of a Poisson sum.

Proposition 9 (Campbell formula)**.** *Assume that ρ has finite α-moment. Let $\{s_j\}_{j\geq 1}$ be the Poisson process in* (48)*, ordered by decreasing size. Independently draw $g_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $\boldsymbol{v}_j \overset{\text{i.i.d.}}{\sim} \rho$. The series $\sum_j \sqrt{s_j} g_j \boldsymbol{v}_j$ converges almost surely and*

$$\mathbb{E}\exp\left(i\sum_j \sqrt{s_j} g_j \boldsymbol{k}^\top \boldsymbol{v}_j\right) = \exp\left\{-d_\alpha \int_0^\infty \mathbb{E}_{g,\boldsymbol{v}}\big[1-\cos(\sqrt{s} g \boldsymbol{k}^\top \boldsymbol{v})\big] s^{-1-\alpha/2}\,\mathrm{d}s\right\} \tag{122}$$

$$= \exp\left\{-\int \rho(\mathrm{d}\boldsymbol{v}) \left|\boldsymbol{k}^\top \boldsymbol{v}\right|^\alpha\right\}. \tag{123}$$

Proof. First recall the Poisson product identity, including its short proof. For disjoint sets B_r of finite intensity $\nu(B_r)$, their counts N_r are independent Poisson variables. Thus a function equal to f_r on B_r and to one elsewhere satisfies

$$\mathbb{E}\prod_j f(X_j) = \prod_r \mathbb{E} f_r^{N_r} = \prod_r \sum_{k=0}^\infty e^{-\nu(B_r)} \frac{\nu(B_r)^k f_r^k}{k!} = \exp\left\{\int (f-1)\,\mathrm{d}\nu\right\}. \tag{124}$$

Approximation by simple functions gives this identity for $f = e^{-h}$, $h \geq 0$, by monotone convergence, and for other integrands whenever both sides and the approximation are integrable. Independent marks are included by replacing ν with intensity times mark distribution. The same simple-function calculation gives $\mathbb{E}\sum_j h(X_j) = \int h\,\mathrm{d}\nu$ for $h \geq 0$.

For any independent scalar or vector mark c, substitution $r = s\|c\|^2$ gives

$$d_\alpha \int_0^\infty \mathbb{E}[1 \wedge s\|c\|^2] s^{-1-\alpha/2}\,\mathrm{d}s = d_\alpha \mathbb{E}\|c\|^\alpha \left[\int_0^1 r^{-\alpha/2}\,\mathrm{d}r + \int_1^\infty r^{-1-\alpha/2}\,\mathrm{d}r\right] \tag{125}$$

$$= \frac{4 d_\alpha}{\alpha(2-\alpha)} \mathbb{E}\|c\|^\alpha. \tag{126}$$

If this quantity is finite, there are almost surely only finitely many terms with $s_j\|c_j\|^2 > 1$, and the sum of the remaining terms has finite expectation. Consequently, $\sum_j s_j\|c_j\|^2 < \infty$ almost surely. Taking $c_j = \boldsymbol{v}_j$, the partial Gaussian sums, conditional on $\{s_j, \boldsymbol{v}_j\}$, form an L^2-bounded martingale and converge almost surely and in conditional L^2. Their characteristic function is

$$\mathbb{E}\exp\left(i\sum_j \sqrt{s_j} g_j \boldsymbol{k}^\top \boldsymbol{v}_j\right) = \mathbb{E}\exp\left(-\frac{1}{2}\sum_j s_j (\boldsymbol{k}^\top \boldsymbol{v}_j)^2\right) \tag{127}$$

$$= \exp\left\{d_\alpha \int_0^\infty \mathbb{E}_{\boldsymbol{v}}\left[e^{-s(\boldsymbol{k}^\top \boldsymbol{v})^2/2} - 1\right] s^{-1-\alpha/2}\,\mathrm{d}s\right\}. \tag{128}$$

Here the second equality is (124). Using $\mathbb{E}_g e^{i\sqrt{s}gz} = e^{-sz^2/2}$ and then (121) proves (123). □

Draw the Poisson points and Gaussian marks in (48), independent objectives $\mathcal{V}_{j,\beta} \sim \Pi_\beta$, and $\epsilon \sim p_\epsilon$. For fixed values of these objects, draw independently over j and a

$$U_j^a \mid \mathcal{V}_{j,\beta} \sim \mu_{\mathcal{V}_{j,\beta}}, \tag{129}$$

where $\mu_{\mathcal{V}}$ is the Gibbs measure defined in (105). Below, $\mathbb{E}_U$ averages only these Gibbs draws, while the outer expectation averages the objectives, Poisson points, Gaussian marks, and noise. Substituting (123) into (88) and using (85) gives

$$\mathcal{I}_n[\rho] = \mathbb{E}\mathbb{E}_U \prod_{a=1}^{n} \left[\int_{\mathbb{R}} \frac{\mathrm{d}k_a}{2\pi}\, \widetilde{g}_\beta(k_a) e^{ik_a(\sum_j \sqrt{s_j} g_j U_j^a - \epsilon)} \right] \tag{130}$$

$$= \mathbb{E} \prod_{a=1}^{n} \mathbb{E}_{U^a} e^{-\beta \ell(\sum_j \sqrt{s_j} g_j U_j^a - \epsilon)} \tag{131}$$

$$= \mathbb{E}\left[\left(\mathbb{E}_U e^{-\beta \ell(\sum_j \sqrt{s_j} g_j U_j - \epsilon)} \right)^n \right]. \tag{132}$$

Here the product separates because the replicas are independent conditional on the same objectives and marks.

We next transform the prefactor in (102). Symmetry of g and (121) give

$$-\left|\boldsymbol{k}^\top \boldsymbol{u}\right|^\alpha = d_\alpha \int_0^\infty \mathbb{E}_g[\cos(\sqrt{s} g \boldsymbol{k}^\top \boldsymbol{u}) - 1] s^{-1-\alpha/2}\, \mathrm{d}s \tag{133}$$

$$= d_\alpha \int_0^\infty \mathbb{E}_g[e^{i\sqrt{s} g \sum_a k_a u^a} - 1] s^{-1-\alpha/2}\, \mathrm{d}s. \tag{134}$$

The additional exponential shifts each Fourier integral explicitly as

$$\int_{\mathbb{R}} \frac{\mathrm{d}k_a}{2\pi}\, \widetilde{g}_\beta(k_a) e^{ik_a(\sum_j \sqrt{s_j} g_j U_j^a - \epsilon)} e^{ik_a \sqrt{s} g u^a} = g_\beta\left(\sqrt{s} g u^a + \sum_j \sqrt{s_j} g_j U_j^a - \epsilon \right) \tag{135}$$

$$= e^{-\beta \ell(\sqrt{s} g u^a + \sum_j \sqrt{s_j} g_j U_j^a - \epsilon)}. \tag{136}$$

Thus (102), (132), and (136) yield

$$\hat{\rho}(\boldsymbol{u}) = \frac{\kappa d_\alpha}{\mathcal{I}_n[\rho]} \int_0^\infty s^{-1-\alpha/2}\, \mathrm{d}s\, \mathbb{E}\left[\prod_{a=1}^{n} \mathbb{E}_{U^a} e^{-\beta \ell(\sqrt{s} g u^a + \sum_j \sqrt{s_j} g_j U_j^a - \epsilon)} \right. \tag{137}$$

$$\left. - \left(\mathbb{E}_U e^{-\beta \ell(\sum_j \sqrt{s_j} g_j U_j - \epsilon)} \right)^n \right] \tag{138}$$

$$= \frac{\kappa d_\alpha}{\mathcal{I}_n[\rho]} \int_0^\infty s^{-1-\alpha/2}\, \mathrm{d}s\, \mathbb{E}\left[\left(\mathbb{E}_U e^{-\beta \ell(\sum_j \sqrt{s_j} g_j U_j - \epsilon)} \right)^n \right. \tag{139}$$

$$\left. \times \left\{ \prod_{a=1}^{n} \frac{\mathbb{E}_{U^a} e^{-\beta \ell(\sqrt{s} g u^a + \sum_j \sqrt{s_j} g_j U_j^a - \epsilon)}}{\mathbb{E}_{U^a} e^{-\beta \ell(\sum_j \sqrt{s_j} g_j U_j^a - \epsilon)}} - 1 \right\} \right]. \tag{140}$$

The $\hat{\rho}$-stationarity equation therefore contains this ratio directly. We define the normalized one-factor message by

$$\hat{\mathcal{W}}_{s,g,\beta}(u) := -\frac{1}{\beta} \log \frac{\mathbb{E}_U e^{-\beta \ell(\sqrt{s} g u + \sum_j \sqrt{s_j} g_j U_j - \epsilon)}}{\mathbb{E}_U e^{-\beta \ell(\sum_j \sqrt{s_j} g_j U_j - \epsilon)}} \tag{141}$$

$$= -\frac{1}{\beta} \log \mathbb{E}_U e^{-\beta \ell(\sqrt{s} g u + \sum_j \sqrt{s_j} g_j U_j - \epsilon)} + \frac{1}{\beta} \log \mathbb{E}_U e^{-\beta \ell(\sum_j \sqrt{s_j} g_j U_j - \epsilon)}. \tag{142}$$

Setting $u = 0$ in (142) gives

$$\hat{\mathcal{W}}_{s,g,\beta}(0) = 0. \tag{143}$$

Substituting (142) into (140), and using (132) for $\mathcal{I}_n[\rho]$, gives

$$\hat{\rho}(\boldsymbol{u}) = \kappa d_\alpha \int_0^\infty \frac{\mathbb{E}\left[\left(\mathbb{E}_U e^{-\beta \ell(\sum_j \sqrt{s_j} g_j U_j - \epsilon)} \right)^n \left(e^{-\beta \sum_a \hat{\mathcal{W}}_{s,g,\beta}(u^a)} - 1 \right) \right]}{\mathbb{E}\left[\left(\mathbb{E}_U e^{-\beta \ell(\sum_j \sqrt{s_j} g_j U_j - \epsilon)} \right)^n \right]} s^{-1-\alpha/2}\, \mathrm{d}s. \tag{144}$$

The replica weight tends to one:

$$\Big(\mathbb{E}_U e^{-\beta\ell(\sum_j \sqrt{s_j} g_j U_j - \epsilon)}\Big)^n \xrightarrow{n\to 0} 1. \tag{145}$$

After this limit, (144) becomes

$$\hat{\rho}_{\mathrm{RS}}(\boldsymbol{u}) = \kappa d_\alpha \int_0^\infty \mathbb{E}\Big[e^{-\beta\sum_a \hat{\mathcal{W}}_{s,g,\beta}(u^a)} - 1\Big] s^{-1-\alpha/2}\,\mathrm{d}s. \tag{146}$$

Exponentiating (146) and applying (124) with intensity $\kappa d_\alpha s^{-1-\alpha/2}\,\mathrm{d}s$ gives

$$e^{\hat{\rho}_{\mathrm{RS}}(\boldsymbol{u})} = \exp\Big\{\kappa d_\alpha \int_0^\infty \mathbb{E}\Big[e^{-\beta\sum_a \hat{\mathcal{W}}_{s,g,\beta}(u^a)} - 1\Big] s^{-1-\alpha/2}\,\mathrm{d}s\Big\} \tag{147}$$

$$= \mathbb{E}\prod_l e^{-\beta\sum_a \hat{\mathcal{W}}^{(l)}_{\hat{s}_l,\hat{g}_l,\beta}(u^a)} \tag{148}$$

$$= \mathbb{E}e^{-\beta\sum_a\sum_l \hat{\mathcal{W}}^{(l)}_{\hat{s}_l,\hat{g}_l,\beta}(u^a)}. \tag{149}$$

Each l uses an independent copy of the sample environment in (142). The aggregate conjugate potential appearing in (149) is therefore

$$\hat{\mathcal{V}}_\beta(u) := \sum_l \hat{\mathcal{W}}^{(l)}_{\hat{s}_l,\hat{g}_l,\beta}(u), \qquad \hat{\mathcal{V}}_\beta(0) = \sum_l \hat{\mathcal{W}}^{(l)}_{\hat{s}_l,\hat{g}_l,\beta}(0) = 0. \tag{150}$$

Substituting (150) into (149) gives

$$e^{\hat{\rho}_{\mathrm{RS}}(\boldsymbol{u})} = \mathbb{E}\prod_{a=1}^n e^{-\beta\hat{\mathcal{V}}_\beta(u^a)} \tag{151}$$

$$= \int \mathrm{Law}(\hat{\mathcal{V}}_\beta)\Big(\mathrm{d}\hat{\mathcal{V}}\Big)\prod_{a=1}^n e^{-\beta\hat{\mathcal{V}}(u^a)}. \tag{152}$$

Comparing (152) with (109) identifies the law of the resulting random function. For each outer Poisson point $\hat{s}_l$ let $\{(s_{lj}, g_{lj}, \mathcal{V}_{lj,\beta}, U_{lj})\}_{j\ge1}, \epsilon_l$ be an independent copy of the sample environment used in (142). Thus $\mathcal{V}_{lj,\beta}$ has law Π_β, and, conditional on it, U_{lj} has Gibbs law $\mu_{\mathcal{V}_{lj,\beta}}$. With $\mathbb{E}_{U_l}$ denoting expectation only over $\{U_{lj}\}_{j\ge1}$, the complete update is

$$\begin{aligned}
\mathcal{V}_{lj,\beta} &\overset{\text{i.i.d.}}{\sim} \Pi_\beta,\\
U_{lj} \mid \mathcal{V}_{lj,\beta} &\sim \mu_{\mathcal{V}_{lj,\beta}},\\
\hat{\mathcal{V}}_\beta(u) &:= -\frac{1}{\beta}\sum_l \log\frac{\mathbb{E}_{U_l}\exp\Big\{-\beta\ell\Big(\sqrt{\hat{s}_l}\hat{g}_l u + \sum_j \sqrt{s_{lj}} g_{lj} U_{lj} - \epsilon_l\Big)\Big\}}{\mathbb{E}_{U_l}\exp\Big\{-\beta\ell\Big(\sum_j \sqrt{s_{lj}} g_{lj} U_{lj} - \epsilon_l\Big)\Big\}},\\
\mathrm{Law}\Big(\hat{\mathcal{V}}_\beta\Big) &= \hat{\Pi}_\beta.
\end{aligned} \tag{153}$$

Together, (120) and (153) are the two finite-temperature RS updates:

$$\hat{\Pi}_\beta \xrightarrow{(120)} \Pi_\beta, \tag{154}$$

$$\Pi_\beta \xrightarrow{(153)} \hat{\Pi}_\beta. \tag{155}$$

C.3 Zero temperature: from coefficient minimizers to one sample

Recall that the finite-temperature fixed point (155) is generated explicitly by

$$\hat{\mathcal{V}}_\beta(u) := \sum_l \hat{\mathcal{W}}^{(l)}_{\hat{s}_l,\hat{g}_l,\beta}(u), \qquad \mathrm{Law}(\hat{\mathcal{V}}_\beta) = \hat{\Pi}_\beta, \tag{156}$$

$$\mathcal{V}_\beta(u) := \lambda[\gamma(w^* + u) - \gamma(w^*)] + \hat{\mathcal{V}}_\beta(u), \qquad \mathrm{Law}(\mathcal{V}_\beta) = \Pi_\beta. \tag{157}$$

Here the one-factor message $\hat{\mathcal{W}}_{s,g,\beta}$ is the logarithmic Gibbs ratio in (142). We now take $\beta\to\infty$ in (157) and write

$$\Pi_\beta \longrightarrow \Pi, \qquad \hat{\Pi}_\beta \longrightarrow \hat{\Pi}. \tag{158}$$

For each independent $\mathcal{V}_j \sim \Pi$, define

$$u_{j,*}(h) := \operatorname*{argmin}_u\{\mathcal{V}_j(u) - hu\}. \tag{159}$$

The $\Pi_\beta \to \hat{\Pi}_\beta$ update as $\beta \to \infty$. The finite-temperature update (153) draws every U_{lj} from the normalized Gibbs measure $\mu_{\mathcal{V}_{l_j,\beta}}$ defined in (105). We must therefore begin by taking the zero-temperature limit of this normalized density. For one generic objective $\mathcal{V}_{j,\beta}$, its negative logarithm divided by β is

$$-\frac{1}{\beta}\log\frac{e^{-\beta\mathcal{V}_{j,\beta}(u)}}{\int_{\mathbb{R}} e^{-\beta\mathcal{V}_{j,\beta}(v)}\,\mathrm{d}v} = \mathcal{V}_{j,\beta}(u) + \frac{1}{\beta}\log\int_{\mathbb{R}} e^{-\beta\mathcal{V}_{j,\beta}(v)}\,\mathrm{d}v \tag{160}$$

$$\xrightarrow{\beta\to\infty} \mathcal{V}_j(u) - \min_v \mathcal{V}_j(v) = \mathcal{V}_j(u) - \mathcal{V}_j(u_{j,*}(0)). \tag{161}$$

The first equality in (161) is the negative logarithm of (105). Its normalizing integral contributes $-\min_v \mathcal{V}_j(v)$ in the limit, and (159) with $h = 0$ gives $\min_v \mathcal{V}_j(v) = \mathcal{V}_j(u_{j,*}(0))$. Thus the normalized Gibbs factor contributes precisely the centered objective appearing on the last line of (161). The two logarithms in (142) therefore give

$$\hat{\mathcal{W}}_{s,g}(u) = \min_{\{u_j\}}\left\{\ell\left(\sqrt{s}gu + \sum_j \sqrt{s_j}g_j u_j - \epsilon\right) + \sum_j [\mathcal{V}_j(u_j) - \mathcal{V}_j(u_{j,*}(0))]\right\} \tag{162}$$

$$- \min_{\{u_j\}}\left\{\ell\left(\sum_j \sqrt{s_j}g_j u_j - \epsilon\right) + \sum_j [\mathcal{V}_j(u_j) - \mathcal{V}_j(u_{j,*}(0))]\right\}. \tag{163}$$

Taking $\beta \to \infty$ in (150) gives

$$\hat{\mathcal{V}}(u) := \sum_l \hat{\mathcal{W}}^{(l)}_{\hat{s}_l,\hat{g}_l}(u), \qquad \mathrm{Law}(\hat{\mathcal{V}}) = \hat{\Pi}. \tag{164}$$

We next evaluate the minimizations in (163), use the result in (164).

Let $m > 0$ be the strong-convexity constant of $\lambda\gamma$. The sample contributions are convex, as the scalar representation below will show, so each $\mathcal{V}_j$ has the same lower bound on its curvature. The optimality condition for (159) and strong convexity give

$$h \in \partial\mathcal{V}_j(u_{j,*}(h)), \tag{165}$$

$$m|u_{j,*}(h) - u_{j,*}(\tilde{h})|^2 \le (h-\tilde{h})[u_{j,*}(h) - u_{j,*}(\tilde{h})] \le |h-\tilde{h}|\,|u_{j,*}(h) - u_{j,*}(\tilde{h})|, \tag{166}$$

$$|u_{j,*}(h) - u_{j,*}(\tilde{h})| \le \frac{|h-\tilde{h}|}{m}, \tag{167}$$

$$\mathcal{V}_j(u) - \mathcal{V}_j(u_{j,*}(0)) \ge \frac{m}{2}|u - u_{j,*}(0)|^2. \tag{168}$$

In particular, $u_{j,*}$ is nondecreasing and Lipschitz.

The coefficient response $\mathcal{R}$ is obtained by perturbing each objective by $\sqrt{s_j}g_j p u$ and subtracting its value and slope at $p = 0$:

$$\mathcal{R}(p) := \sum_j \Big[\min_u\{\mathcal{V}_j(u) + \sqrt{s_j}g_j p u\} - \mathcal{V}_j(u_{j,*}(0)) - \sqrt{s_j}g_j p\, u_{j,*}(0)\Big] \tag{169}$$

$$= \sum_j \min_u\big\{\mathcal{V}_j(u) - \mathcal{V}_j(u_{j,*}(0)) + \sqrt{s_j}g_j p[u - u_{j,*}(0)]\big\}. \tag{170}$$

Evaluation at $u = u_{j,*}(0)$, followed by (168) and completion of the square, proves

$$0 \ge \min_u\big\{\mathcal{V}_j(u) - \mathcal{V}_j(u_{j,*}(0)) + \sqrt{s_j}g_j p[u - u_{j,*}(0)]\big\} \tag{171}$$

$$\ge \min_u\Big\{\frac{m}{2}[u - u_{j,*}(0)]^2 + \sqrt{s_j}g_j p[u - u_{j,*}(0)]\Big\} \tag{172}$$

$$= \min_u \frac{m}{2}\left[u - u_{j,*}(0) + \frac{\sqrt{s_j}g_j p}{m}\right]^2 - \frac{s_j g_j^2 p^2}{2m} = -\frac{s_j g_j^2 p^2}{2m}. \tag{173}$$

To differentiate the minimum, evaluate the objective at the minimizer for $p+\delta$. For $\delta > 0$, this gives the two bounds

$$\sqrt{s_j}g_j u_{j,*}(-\sqrt{s_j}g_j(p+\delta)) \le \frac{1}{\delta}\Big[\min_u\{\mathcal{V}_j(u) + \sqrt{s_j}g_j(p+\delta)u\} - \min_u\{\mathcal{V}_j(u) + \sqrt{s_j}g_j p u\}\Big] \tag{174}$$

$$\le \sqrt{s_j}g_j u_{j,*}(-\sqrt{s_j}g_j p). \tag{175}$$

Both bounds have the same limit by (167), so

$$\frac{\mathrm{d}}{\mathrm{d}p}\min_u\{\mathcal{V}_j(u)+\sqrt{s_j}g_jpu\}=\sqrt{s_j}g_ju_{j,*}(-\sqrt{s_j}g_jp), \tag{176}$$

$$\mathcal{R}'(p)=\sum_j\left[\frac{\mathrm{d}}{\mathrm{d}p}\min_u\{\mathcal{V}_j(u)+\sqrt{s_j}g_jpu\}-\sqrt{s_j}g_ju_{j,*}(0)\right] \tag{177}$$

$$=\sum_j\sqrt{s_j}g_j[u_{j,*}(-\sqrt{s_j}g_jp)-u_{j,*}(0)],\qquad \mathcal{R}(0)=\mathcal{R}'(0)=0. \tag{178}$$

This is (55). Each minimum in (170) is concave in p, since it is an infimum of affine functions, and the subtracted terms are affine. Thus $\mathcal{R}$ is concave.

The series and its derivative converge locally uniformly because

$$\mathbb{E}\sum_{j:s_j\le 1}s_jg_j^2=d_\alpha\int_0^1 s\,s^{-1-\alpha/2}\,\mathrm{d}s=\frac{2d_\alpha}{2-\alpha}<\infty,\quad \mathbb{E}\#\{j:s_j>1\}=\frac{2d_\alpha}{\alpha}<\infty, \tag{179}$$

$$\sum_j\sup_{|p|\le H}|\sqrt{s_j}g_j[u_{j,*}(-\sqrt{s_j}g_jp)-u_{j,*}(0)]|\le\frac{H}{m}\sum_j s_jg_j^2<\infty. \tag{180}$$

The same sum bounds the absolute values of the centered minima in (173) by $H^2\sum_j s_jg_j^2/(2m)$. This justifies the termwise differentiation in (178).

The linear terms subtracted in (170) contain the uncentered Poisson series marked by $u_{j,*}(0)$. At this point we restrict attention to fixed-point solutions satisfying

$$\mathbb{E}_{\mathcal{V}\sim\Pi}|u_*(0)|^\alpha<\infty, \tag{181}$$

where $u_*(0)$ denotes a generic copy of the minimizer in (159). Applying (126) with mark $u_*(0)$ gives

$$\mathbb{E}\sum_j\big[1\wedge s_ju_{j,*}(0)^2\big]=\frac{4d_\alpha}{\alpha(2-\alpha)}\mathbb{E}|u_*(0)|^\alpha<\infty. \tag{182}$$

Hence only finitely many terms satisfy $s_ju_{j,*}(0)^2>1$, while the remaining terms have a finite sum, so

$$\sum_j s_ju_{j,*}(0)^2<\infty\qquad\text{almost surely.} \tag{183}$$

The scale and its standardized Gaussian field are therefore well defined by

$$Q:=\sum_j s_ju_{j,*}(0)^2,\qquad\qquad \xi:=\frac{1}{\sqrt{Q}}\sum_j\sqrt{s_j}g_ju_{j,*}(0). \tag{184}$$

Conditional on $\{s_j,\mathcal{V}_j\}$, Proposition 9 gives a convergent Gaussian series and, for $Q>0$,

$$\mathbb{E}_g[e^{it\xi}\mid\{s_j,\mathcal{V}_j\}]=\prod_j\mathbb{E}_{g_j}e^{it\sqrt{s_j}g_ju_{j,*}(0)/\sqrt{Q}}=\prod_j e^{-t^2s_ju_{j,*}(0)^2/(2Q)} \tag{185}$$

$$=\exp\left(-\frac{t^2}{2Q}\sum_j s_ju_{j,*}(0)^2\right)=e^{-t^2/2}. \tag{186}$$

On $Q=0$, take an independent standard Gaussian for ξ. The conditional characteristic function (186) is independent of Q, so $\xi\sim\mathcal{N}(0,1)$ is independent of Q. The variables ξ and $\mathcal{R}$ are dependent because both are functions of the same g_j. Equations (178) and (184) therefore define

$$\mathcal{P}:=\mathrm{Law}(Q,\xi,\mathcal{R}). \tag{187}$$

For every configuration of finite centered energy in (163), its preactivation is also well defined:

$$\sum_j |\sqrt{s_j} g_j [u_j - u_{j,*}(0)]| \le \left(\sum_j s_j g_j^2 \right)^{1/2} \left(\sum_j |u_j - u_{j,*}(0)|^2 \right)^{1/2} < \infty, \tag{188}$$

$$\sum_j \sqrt{s_j} g_j u_j = \sum_j \sqrt{s_j} g_j u_{j,*}(0) + \sum_j \sqrt{s_j} g_j [u_j - u_{j,*}(0)] \tag{189}$$

$$= \sqrt{Q}\, \xi + \sum_j \sqrt{s_j} g_j [u_j - u_{j,*}(0)]. \tag{190}$$

The finiteness in the first line follows from (168) and (179).

Insert $\ell(r) = \sup_{p \in K_\ell} \{pr - \ell^*(p)\}$ into the minimum in (163), with a general shift h. Then use (170) and (184):

$$\min_{\{u_j\}} \left\{ \ell \left(h + \sum_j \sqrt{s_j} g_j u_j - \epsilon \right) + \sum_j [\mathcal{V}_j(u_j) - \mathcal{V}_j(u_{j,*}(0))] \right\} \tag{191}$$

$$= \min_{\{u_j\}} \sup_{p \in K_\ell} \left\{ (h - \epsilon) p - \ell^*(p) + \sum_j [\mathcal{V}_j(u_j) + \sqrt{s_j} g_j p u_j - \mathcal{V}_j(u_{j,*}(0))] \right\} \tag{192}$$

$$= \sup_{p \in K_\ell} \left\{ (h - \epsilon) p - \ell^*(p) + \sum_j \Big[\min_u \{\mathcal{V}_j(u) + \sqrt{s_j} g_j p u\} - \mathcal{V}_j(u_{j,*}(0)) \Big] \right\} \tag{193}$$

$$= \sup_{p \in K_\ell} \left\{ (h - \epsilon) p - \ell^*(p) + \mathcal{R}(p) + p \sum_j \sqrt{s_j} g_j u_{j,*}(0) \right\} \tag{194}$$

$$= \sup_{p \in K_\ell} \{ (\sqrt{Q}\, \xi - \epsilon + h) p - \ell^*(p) + \mathcal{R}(p) \} = (\ell^* - \mathcal{R})^* (\sqrt{Q}\, \xi - \epsilon + h). \tag{195}$$

The exchange of the minimum and supremum can be checked at the optimum. The primal optimality conditions give a common p with

$$p \in \partial \ell \left(h + \sum_j \sqrt{s_j} g_j u_j - \epsilon \right), \qquad -\sqrt{s_j} g_j p \in \partial \mathcal{V}_j(u_j), \qquad u_j = u_{j,*}(-\sqrt{s_j} g_j p). \tag{196}$$

At these values, both parts of the primal objective attain their Fenchel bounds:

$$\ell \left(h + \sum_j \sqrt{s_j} g_j u_j - \epsilon \right) = (h - \epsilon) p + p \sum_j \sqrt{s_j} g_j u_j - \ell^*(p), \tag{197}$$

$$\mathcal{V}_j(u_j) - \mathcal{V}_j(u_{j,*}(0)) = \min_u \{\mathcal{V}_j(u) + \sqrt{s_j} g_j p u\} - \mathcal{V}_j(u_{j,*}(0)) - \sqrt{s_j} g_j p u_j. \tag{198}$$

Adding (198) cancels $p \sum_j \sqrt{s_j} g_j u_j$ and gives exactly the third line of (195).

To obtain a minimum over the residual v, apply Fenchel's inequality to the two terms below:

$$\ell(v) + (-\mathcal{R})^* (\sqrt{Q}\, \xi - \epsilon + h - v) \ge pv - \ell^*(p) + p(\sqrt{Q}\, \xi - \epsilon + h - v) + \mathcal{R}(p) \tag{199}$$

$$= (\sqrt{Q}\, \xi - \epsilon + h) p - \ell^*(p) + \mathcal{R}(p). \tag{200}$$

For the minimizing coordinates in (196), the residual satisfies

$$v = h - \epsilon + \sum_j \sqrt{s_j} g_j u_{j,*}(-\sqrt{s_j} g_j p) \tag{201}$$

$$= \sqrt{Q}\, \xi - \epsilon + h + \sum_j \sqrt{s_j} g_j [u_{j,*}(-\sqrt{s_j} g_j p) - u_{j,*}(0)] \tag{202}$$

$$= \sqrt{Q}\, \xi - \epsilon + h + \mathcal{R}'(p), \qquad p \in \partial \ell(v). \tag{203}$$

Thus both inequalities in (200) become equalities:

$$\ell(v) = pv - \ell^*(p), \tag{204}$$

$$(-\mathcal{R})^*(\sqrt{Q}\,\xi - \epsilon + h - v) = p(\sqrt{Q}\,\xi - \epsilon + h - v) + \mathcal{R}(p), \tag{205}$$

$$\ell(v) + (-\mathcal{R})^*(\sqrt{Q}\,\xi - \epsilon + h - v) = (\sqrt{Q}\,\xi - \epsilon + h)p - \ell^*(p) + \mathcal{R}(p) \tag{206}$$

$$= (\ell^* - \mathcal{R})^*(\sqrt{Q}\,\xi - \epsilon + h). \tag{207}$$

We have therefore proved the scalar representation

$$(\ell^* - \mathcal{R})^*(\sqrt{Q}\,\xi - \epsilon + h) = \min_v\{\ell(v) + (-\mathcal{R})^*(\sqrt{Q}\,\xi - \epsilon + h - v)\}, \tag{208}$$

$$v_*(h) \in \operatorname*{argmin}_v\{\ell(v) + (-\mathcal{R})^*(\sqrt{Q}\,\xi - \epsilon + h - v)\}. \tag{209}$$

This is (59). For the maximizing p in (195), evaluating the perturbed supremum at that same p gives

$$(\ell^* - \mathcal{R})^*(\sqrt{Q}\,\xi - \epsilon + h + \delta) \geq (\sqrt{Q}\,\xi - \epsilon + h + \delta)p - \ell^*(p) + \mathcal{R}(p) \tag{210}$$

$$= (\ell^* - \mathcal{R})^*(\sqrt{Q}\,\xi - \epsilon + h) + p\delta. \tag{211}$$

Hence the envelope subgradient is the common loss subgradient in (203):

$$\frac{\mathrm{d}}{\mathrm{d}h}(\ell^* - \mathcal{R})^*(\sqrt{Q}\,\xi - \epsilon + h) = p = \ell'(v_*(h)). \tag{212}$$

At nonsmooth points, the derivatives denote this common subgradient selection.

For each hatted Poisson point, draw an independent $(Q_l, \xi_l, \mathcal{R}_l) \sim \mathcal{P}$ and $\epsilon_l \sim p_\epsilon$. Equations (163), (208), and (212) give the entire sample contribution as

$$\hat{\mathcal{W}}^{(l)}_{\hat{s}_l,\hat{g}_l}(u) = (\ell^* - \mathcal{R}_l)^*(\sqrt{Q_l}\xi_l - \epsilon_l + \sqrt{\hat{s}_l}\hat{g}_l u) - (\ell^* - \mathcal{R}_l)^*(\sqrt{Q_l}\xi_l - \epsilon_l) \tag{213}$$

$$= \int_0^u \sqrt{\hat{s}_l}\hat{g}_l \ell'(v_{l,*}(\sqrt{\hat{s}_l}\hat{g}_l t))\,\mathrm{d}t \tag{214}$$

$$= \sqrt{\hat{s}_l}\hat{g}_l \ell'(v_{l,*}(0))u + \int_0^u \sqrt{\hat{s}_l}\hat{g}_l[\ell'(v_{l,*}(\sqrt{\hat{s}_l}\hat{g}_l t)) - \ell'(v_{l,*}(0))]\,\mathrm{d}t. \tag{215}$$

The centered terms define the coefficient-side response directly:

$$\hat{\mathcal{R}}(u) := \sum_l \Big[\hat{\mathcal{W}}^{(l)}_{\hat{s}_l,\hat{g}_l}(u) - \sqrt{\hat{s}_l}\hat{g}_l \ell'(v_{l,*}(0))u\Big] \tag{216}$$

$$= \sum_l \int_0^u \sqrt{\hat{s}_l}\hat{g}_l[\ell'(v_{l,*}(\sqrt{\hat{s}_l}\hat{g}_l t)) - \ell'(v_{l,*}(0))]\,\mathrm{d}t, \tag{217}$$

$$\hat{\mathcal{R}}'(u) = \sum_l \sqrt{\hat{s}_l}\hat{g}_l[\ell'(v_{l,*}(\sqrt{\hat{s}_l}\hat{g}_l u)) - \ell'(v_{l,*}(0))]. \tag{218}$$

This is (62). The conjugate $(\ell^* - \mathcal{R}_l)^*$ is convex, so $h \mapsto \ell'(v_{l,*}(h))$ is nondecreasing by (212). Each centered term in (218) is therefore convex, vanishes at zero, and has zero as a subgradient there.

C.3.1 Convergence of the sample response

We verify the termwise integration and differentiation in (218). For globally L_ℓ-Lipschitz loss, the slope in (212) lies in $[-L_\ell, L_\ell]$. Indeed, for $|p| > L_\ell$,

$$\ell^*(p) = \sup_r\{pr - \ell(r)\} \geq \sup_{r:\, pr \geq 0}\{(|p| - L_\ell)|r| - \ell(0)\} = \infty, \qquad K_\ell \subset [-L_\ell, L_\ell]. \tag{219}$$

Condition on $(Q_l, \xi_l, \mathcal{R}_l)$. For $t \geq 0$, change variables from ϵ_l to $a = \sqrt{Q_l}\xi_l - \epsilon_l$ in the noise average:

$$\mathbb{E}_{\epsilon_l}|\ell'(v_{l,*}(t)) - \ell'(v_{l,*}(0))| \tag{220}$$

$$= \int_{\mathbb{R}} p_\epsilon(\sqrt{Q_l}\xi_l - a)[(\ell^* - \mathcal{R}_l)^{*\prime}(a+t) - (\ell^* - \mathcal{R}_l)^{*\prime}(a)]\,\mathrm{d}a \tag{221}$$

$$\leq \sup_x p_\epsilon(x) \int_{\mathbb{R}} [(\ell^* - \mathcal{R}_l)^{*\prime}(a+t) - (\ell^* - \mathcal{R}_l)^{*\prime}(a)]\,\mathrm{d}a \tag{222}$$

$$= t \sup_x p_\epsilon(x) \left[\lim_{a\to\infty} (\ell^* - \mathcal{R}_l)^{*\prime}(a) - \lim_{a\to-\infty} (\ell^* - \mathcal{R}_l)^{*\prime}(a)\right] \leq 2L_\ell \sup_x p_\epsilon(x) t. \tag{223}$$

For the penultimate equality, shifting the first integral cancels the common interval, leaving

$$\int_{-B}^{B} [(\ell^* - \mathcal{R}_l)^{*\prime}(a+t) - (\ell^* - \mathcal{R}_l)^{*\prime}(a)]\,\mathrm{d}a \tag{224}$$

$$= \int_{B}^{B+t} (\ell^* - \mathcal{R}_l)^{*\prime}(a)\,\mathrm{d}a - \int_{-B}^{-B+t} (\ell^* - \mathcal{R}_l)^{*\prime}(a)\,\mathrm{d}a. \tag{225}$$

Taking $B \to \infty$ gives the stated limits; exchanging zero and t gives (223) with $|t|$ for negative t. Monotonicity also bounds every slope increment on $[-H, H]$ by the two endpoint increments, so

$$\mathbb{E}\left[\sup_{|u|\leq H} |\sqrt{\hat{s}_l}\hat{g}_l[\ell'(v_{l,*}(\sqrt{\hat{s}_l}\hat{g}_l u)) - \ell'(v_{l,*}(0))]| \,\middle|\, \hat{s}_l, \hat{g}_l\right] \tag{226}$$

$$\leq \sqrt{\hat{s}_l}|\hat{g}_l|\,\mathbb{E}\left[\ell'(v_{l,*}(\sqrt{\hat{s}_l}|\hat{g}_l|H)) - \ell'(v_{l,*}(-\sqrt{\hat{s}_l}|\hat{g}_l|H)) \,\middle|\, \hat{s}_l, \hat{g}_l\right] \tag{227}$$

$$\leq 4L_\ell \sup_x p_\epsilon(x) H \hat{s}_l \hat{g}_l^2. \tag{228}$$

Integrating over u and summing the points with $\hat{s}_l \leq 1$ yields

$$\mathbb{E} \sum_{l:\hat{s}_l\leq 1} \sup_{|u|\leq H} \left|\int_0^u \sqrt{\hat{s}_l}\hat{g}_l[\ell'(v_{l,*}(\sqrt{\hat{s}_l}\hat{g}_l t)) - \ell'(v_{l,*}(0))]\,\mathrm{d}t\right| \tag{229}$$

$$\leq 4L_\ell \sup_x p_\epsilon(x) H^2 \mathbb{E} \sum_{l:\hat{s}_l\leq 1} \hat{s}_l\hat{g}_l^2 = \frac{8\kappa d_\alpha L_\ell \sup_x p_\epsilon(x) H^2}{2-\alpha} < \infty. \tag{230}$$

There are finitely many points above one, so both the centered values and their selected slopes have absolutely convergent series, locally uniformly. A convex function on the line has at most countably many nondifferentiability points, and the independent noise has a density. Consequently, the baseline slope in (218) is unambiguous almost surely.

For squared loss, $\ell^*(p) = p^2/2$. The optimum in (195) satisfies $\sqrt{Q_l}\xi_l - \epsilon_l + h \in p + \partial(-\mathcal{R}_l)(p)$, with $p = \ell'(v_{l,*}(h)) = v_{l,*}(h)$. Monotonicity of $\partial(-\mathcal{R}_l)$ therefore gives

$$0 \leq [h - \tilde{h} - v_{l,*}(h) + v_{l,*}(\tilde{h})][v_{l,*}(h) - v_{l,*}(\tilde{h})], \tag{231}$$

$$|v_{l,*}(h) - v_{l,*}(\tilde{h})|^2 \leq (h - \tilde{h})[v_{l,*}(h) - v_{l,*}(\tilde{h})], \tag{232}$$

$$|v_{l,*}(h) - v_{l,*}(\tilde{h})| \leq |h - \tilde{h}|. \tag{233}$$

$$0 \leq \int_0^u \sqrt{\hat{s}_l}\hat{g}_l[v_{l,*}(\sqrt{\hat{s}_l}\hat{g}_l t) - v_{l,*}(0)]\,\mathrm{d}t \tag{234}$$

$$\leq \int_0^{|u|} \hat{s}_l\hat{g}_l^2 t\,\mathrm{d}t = \frac{1}{2}\hat{s}_l\hat{g}_l^2 u^2. \tag{235}$$

The hatted version of (179) proves the same convergence for squared loss. Thus (218) defines a convex $\hat{\mathcal{R}}$ with $\hat{\mathcal{R}}(0) = 0$ and $0 \in \partial\hat{\mathcal{R}}(0)$.

C.3.2 THE LINEAR FIELD AND THE FINAL COEFFICIENT PROBLEM

The linear terms in (215) contain the second uncentered Poisson series, now marked by the sample score $\ell'(v_{l,*}(0))$. Since the sample problem (209) and the law $\mathcal{P}$ have now been constructed, we impose the required moment condition here:

$$\mathbb{E}_{(Q,\xi,\mathcal{R})\sim\mathcal{P},\,\epsilon\sim p_\epsilon}\,|\ell'(v_*(0))|^\alpha < \infty. \tag{236}$$

The expectation includes both the random sample problem and its noise. For a globally L_ℓ-Lipschitz loss this condition is automatic because $|\ell'(v_*(0))| \le L_\ell$. For squared loss, $\ell'(v) = v$, it becomes

$$\mathbb{E}\,|v_*(0)|^\alpha < \infty. \tag{237}$$

The marks are independent of $\{\hat{s}_l, \hat{g}_l\}$, so the hatted version of (126) gives

$$\mathbb{E}\sum_l \big[1 \wedge \hat{s}_l \ell'(v_{l,*}(0))^2\big] = \frac{4\kappa d_\alpha}{\alpha(2-\alpha)}\mathbb{E}\,|\ell'(v_*(0))|^\alpha < \infty. \tag{238}$$

Exactly as in (183), this implies

$$\sum_l \hat{s}_l \ell'(v_{l,*}(0))^2 < \infty \qquad \text{almost surely.} \tag{239}$$

Thus the sample-side scale and standardized field are well defined by

$$\hat{Q} := \sum_l \hat{s}_l \ell'(v_{l,*}(0))^2, \qquad \zeta := \frac{1}{\sqrt{\hat{Q}}}\sum_l \sqrt{\hat{s}_l}\hat{g}_l \ell'(v_{l,*}(0)). \tag{240}$$

Conditional on the hatted points and sample problems, the Gaussian calculation is

$$\mathbb{E}_{\hat{g}}[e^{it\zeta} \mid \{\hat{s}_l, Q_l, \xi_l, \mathcal{R}_l, \epsilon_l\}] = \prod_l e^{-t^2 \hat{s}_l \ell'(v_{l,*}(0))^2/(2\hat{Q})} \tag{241}$$

$$= \exp\left(-\frac{t^2}{2\hat{Q}}\sum_l \hat{s}_l \ell'(v_{l,*}(0))^2\right) = e^{-t^2/2}. \tag{242}$$

On $\hat{Q} = 0$, take an independent standard Gaussian for ζ. This proves $\zeta \sim \mathcal{N}(0,1)$, independent of $\hat{Q}$, and defines

$$\hat{\mathcal{P}} := \mathrm{Law}(\hat{Q}, \zeta, \hat{\mathcal{R}}). \tag{243}$$

As on the coefficient side, ζ and $\hat{\mathcal{R}}$ are dependent because both are functions of the same Gaussian marks.

For completeness, the scale distributions follow directly from (124):

$$-\log \mathbb{E}e^{-zQ} = d_\alpha \int_0^\infty \mathbb{E}[1 - e^{-zsu_*(0)^2}] s^{-1-\alpha/2}\,\mathrm{d}s \tag{244}$$

$$= d_\alpha z^{\alpha/2}\mathbb{E}|u_*(0)|^\alpha \int_0^\infty (1-e^{-r}) r^{-1-\alpha/2}\,\mathrm{d}r \tag{245}$$

$$= d_\alpha \frac{\Gamma(1-\alpha/2)}{\alpha/2} z^{\alpha/2}\mathbb{E}|u_*(0)|^\alpha = 2^{\alpha/2} z^{\alpha/2}\mathbb{E}|u_*(0)|^\alpha, \tag{246}$$

$$-\log \mathbb{E}e^{-z\hat{Q}} = \kappa d_\alpha \int_0^\infty \mathbb{E}[1 - e^{-zs\ell'(v_*(0))^2}] s^{-1-\alpha/2}\,\mathrm{d}s \tag{247}$$

$$= \kappa\, 2^{\alpha/2} z^{\alpha/2}\mathbb{E}|\ell'(v_*(0))|^\alpha. \tag{248}$$

Here integration by parts gives the Gamma integral, and the duplication and reflection identities give $d_\alpha \Gamma(1-\alpha/2)/(\alpha/2) = 2^{\alpha/2}$. Thus both scales are positive $\alpha/2$-stable variables, or zero when the corresponding moment vanishes.

Finally, insert (215) into (164), and then use (218) and (240):

$$\hat{\mathcal{V}}(u) = \sum_l \hat{\mathcal{W}}^{(l)}_{\hat{s}_l,\hat{g}_l}(u) \tag{249}$$

$$= u\sum_l \sqrt{\hat{s}_l}\hat{g}_l \ell'(v_{l,*}(0)) + \sum_l \int_0^u \sqrt{\hat{s}_l}\hat{g}_l \Big[\ell'(v_{l,*}(\sqrt{\hat{s}_l}\hat{g}_l t)) - \ell'(v_{l,*}(0))\Big]\,\mathrm{d}t \tag{250}$$

$$= \sqrt{\hat{Q}}\,\zeta u + \hat{\mathcal{R}}(u). \tag{251}$$

The $\hat{\Pi}_\beta \to \Pi_\beta$ update as $\beta \to \infty$. Taking $\beta \to \infty$ in (114) gives

$$\mathcal{V}(u) := \lambda[\gamma(w^* + u) - \gamma(w^*)] + \hat{\mathcal{V}}(u), \qquad \mathrm{Law}(\mathcal{V}) = \Pi. \tag{252}$$

Substituting (251) into (252) gives

$$\begin{aligned}
\mathcal{V}(u) &= \lambda[\gamma(w^* + u) - \gamma(w^*)] + \hat{\mathcal{V}}(u) && (253)\\
&= \lambda[\gamma(w^* + u) - \gamma(w^*)] + \sqrt{\hat{Q}}\,\zeta u + \hat{\mathcal{R}}(u). && (254)\\
u_*(h) &= \operatorname*{argmin}_u\{\mathcal{V}(u) - hu\} && (255)\\
&= \operatorname*{argmin}_u\Big\{\lambda[\gamma(w^* + u) - \gamma(w^*)] + \sqrt{\hat{Q}}\,\zeta u + \hat{\mathcal{R}}(u) - hu\Big\}. && (256)
\end{aligned}$$

Convexity of $\hat{\mathcal{R}}$ establishes the strong convexity used in (168). Equation (256) is (52).

Proposition 10 (Zero-temperature RS saddle point)**.** *The zero-temperature functional saddle is specified by the following two explicit distributional updates. For independent copies $(\hat{Q}_j, \zeta_j, \hat{\mathcal{R}}_j) \sim \hat{\mathcal{P}}$ and $w_j^* \sim p_w$, independent of $\{(s_j, g_j)\}_{j\ge 1}$, define*

$$\begin{aligned}
u_{j,*}(h) &:= \operatorname*{argmin}_u\Big\{\lambda[\gamma(w_j^* + u) - \gamma(w_j^*)] + \sqrt{\hat{Q}_j}\,\zeta_j u + \hat{\mathcal{R}}_j(u) - hu\Big\}, && (257)\\
Q &:= \sum_j s_j u_{j,*}(0)^2, && (258)\\
\xi &:= \frac{1}{\sqrt{Q}}\sum_j \sqrt{s_j}\, g_j u_{j,*}(0), && (259)\\
\mathcal{R}'(h) &:= \sum_j \sqrt{s_j}\, g_j\big[u_{j,*}(-\sqrt{s_j}\, g_j h) - u_{j,*}(0)\big], && (260)\\
\mathcal{R}(0) &:= 0, && (261)\\
\mathrm{Law}(Q, \xi, \mathcal{R}) &= \mathcal{P}. && (262)
\end{aligned}$$

For independent copies $(Q_l, \xi_l, \mathcal{R}_l) \sim \mathcal{P}$ and $\epsilon_l \sim p_\epsilon$, independent of $\{(\hat{s}_l, \hat{g}_l)\}_{l\ge 1}$, define

$$\begin{aligned}
v_{l,*}(h) &\in \operatorname*{argmin}_v\Big\{\ell(v) + (-\mathcal{R}_l)^*(\sqrt{Q_l}\,\xi_l - \epsilon_l + h - v)\Big\}, && (263)\\
\hat{Q} &:= \sum_l \hat{s}_l \ell'(v_{l,*}(0))^2, && (264)\\
\zeta &:= \frac{1}{\sqrt{\hat{Q}}}\sum_l \sqrt{\hat{s}_l}\,\hat{g}_l \ell'(v_{l,*}(0)), && (265)\\
\hat{\mathcal{R}}'(h) &:= \sum_l \sqrt{\hat{s}_l}\,\hat{g}_l\Big[\ell'(v_{l,*}(\sqrt{\hat{s}_l}\,\hat{g}_l h)) - \ell'(v_{l,*}(0))\Big], && (266)\\
\hat{\mathcal{R}}(0) &:= 0, && (267)\\
\mathrm{Law}(\hat{Q}, \zeta, \hat{\mathcal{R}}) &= \hat{\mathcal{P}}. && (268)
\end{aligned}$$

Equivalently, the two displayed updates satisfy the distributional equalities

$$\begin{aligned}
\hat{\mathcal{P}} &\longrightarrow \mathrm{Law}(Q, \xi, \mathcal{R}) = \mathcal{P}, && (269)\\
\mathcal{P} &\longrightarrow \mathrm{Law}(\hat{Q}, \zeta, \hat{\mathcal{R}}) = \hat{\mathcal{P}}. && (270)
\end{aligned}$$

Proof. The conjugate law closure (164) and (251) give

$$\begin{aligned}
\hat{\Pi} &= \mathrm{Law}\left(u \mapsto \sum_l \hat{\mathcal{W}}^{(l)}_{\hat{s}_l, \hat{g}_l}(u)\right) && (271)\\
&= \mathrm{Law}\left(u \mapsto \sqrt{\hat{Q}}\,\zeta u + \hat{\mathcal{R}}(u)\right). && (272)
\end{aligned}$$

Substituting this identity into the coefficient law closure (252), equivalently using (254), gives

$$\Pi = \mathrm{Law}\Big(u \mapsto \lambda[\gamma(w^* + u) - \gamma(w^*)] + \hat{\mathcal{V}}(u)\Big) \tag{273}$$

$$= \mathrm{Law}\bigg(u \mapsto \lambda[\gamma(w^* + u) - \gamma(w^*)] + \sqrt{\hat{Q}}\,\zeta u + \hat{\mathcal{R}}(u)\bigg). \tag{274}$$

Thus the reconstructed coefficient objective has exactly the law of the objectives used in (159). Applying (178) and (184) to independent copies of this objective gives quantities with joint law $\mathcal{P}$; applying (218) and (240) to independent copies from $\mathcal{P}$ gives quantities with joint law $\hat{\mathcal{P}}$. These are (56) and (63). □

C.4 From the RS saddle to observable marginals

The preceding subsection derived the closed RS saddle equations and the two associated one-dimensional problems. At that stage, however, $u_*(0)$ and $v_*(0)$ are only solutions of the effective coefficient and sample problems; their relation to the original high-dimensional estimator has not yet been established. We now show that their distributions are, respectively, the limiting one-coordinate law of the coefficient error and the limiting empirical distribution of the training residual. This identifies the one-dimensional saddle problems with observable distributions of the original estimator.

Recall from (256) and (209) that the two effective solutions at zero external field are

$$u_*(0) = \operatorname*{argmin}_u \bigg\{\lambda[\gamma(w^* + u) - \gamma(w^*)] + \sqrt{\hat{Q}}\,\zeta u + \hat{\mathcal{R}}(u)\bigg\}, \tag{275}$$

$$v_*(0) \in \operatorname*{argmin}_v \Big\{\ell(v) + (-\mathcal{R})^*\big(\sqrt{Q}\,\xi - \epsilon - v\big)\Big\}. \tag{276}$$

We now identify the first random variable in (276) with a coefficient error and the second with a training residual.

For a bounded continuous test function ϕ, the one-coordinate Gibbs marginal in (119) gives

$$\mathbb{E}\int_{\mathbb{R}} \phi(u)\mu_{\mathcal{V}_\beta}(\mathrm{d}u) = \mathbb{E}\frac{\int_{\mathbb{R}} \phi(u)e^{-\beta\mathcal{V}_\beta(u)}\,\mathrm{d}u}{\int_{\mathbb{R}} e^{-\beta\mathcal{V}_\beta(u)}\,\mathrm{d}u} \tag{277}$$

$$\xrightarrow{\beta\to\infty} \mathbb{E}\phi\bigg(\operatorname*{argmin}_u \mathcal{V}(u)\bigg) = \mathbb{E}\phi(u_*(0)). \tag{278}$$

The minimizer is unique by (254). Therefore the coefficient-error conclusion is explicitly

$$\hat{w}_i - w_i^* \xrightarrow{\mathrm{d}} u_{i,*}(0). \tag{279}$$

This is the law stated in (69).

For the training residual, insert the source $J\phi$ into each loss factor. At finite D, this multiplies (74) by the Gibbs expectation

$$Z_{\beta,J} := Z_\beta \left\langle e^{J\sum_\mu \phi(D^{-1/\alpha}\boldsymbol{x}_\mu^\top \boldsymbol{u} - \epsilon_\mu)}\right\rangle_\beta, \tag{280}$$

$$\frac{1}{N}\partial_J \log Z_{\beta,J}\bigg|_{J=0} = \frac{1}{N}\sum_\mu \Big\langle \phi(D^{-1/\alpha}\boldsymbol{x}_\mu^\top \boldsymbol{u} - \epsilon_\mu)\Big\rangle_\beta \tag{281}$$

$$\xrightarrow{\beta\to\infty} \frac{1}{N}\sum_\mu \phi\Big(D^{-1/\alpha}\boldsymbol{x}_\mu^\top(\hat{\boldsymbol{w}} - \boldsymbol{w}^*) - \epsilon_\mu\Big). \tag{282}$$

The same source changes (132) to

$$\mathcal{I}_{n,J}[\rho] = \mathbb{E}\Big[\Big(\mathbb{E}_U e^{-\beta\ell(\sum_j \sqrt{s_j}g_jU_j - \epsilon) + J\phi(\sum_j \sqrt{s_j}g_jU_j - \epsilon)}\Big)^n\Big]. \tag{283}$$

At the saddle, the variations with respect to ρ and $\hat{\rho}$ vanish, so only the explicit source dependence contributes when differentiating (97):

$$\frac{1}{N}\partial_J \mathbb{E}\log Z_{\beta,J}\Big|_{J=0} = \lim_{n\to 0}\frac{1}{\kappa n}\partial_J \mathcal{S}_n[\rho,\hat{\rho};J]\Big|_{J=0} = \lim_{n\to 0}\frac{1}{n}\partial_J \log \mathcal{I}_{n,J}[\rho]\Big|_{J=0} \tag{284}$$

$$= \partial_J \mathbb{E}\log \mathbb{E}_U e^{-\beta\ell(\sum_j \sqrt{s_j}g_jU_j-\epsilon)+J\phi(\sum_j \sqrt{s_j}g_jU_j-\epsilon)}\Big|_{J=0} \tag{285}$$

$$= \mathbb{E}\frac{\mathbb{E}_U\left[\phi(\sum_j \sqrt{s_j}g_jU_j-\epsilon)e^{-\beta\ell(\sum_j \sqrt{s_j}g_jU_j-\epsilon)}\right]}{\mathbb{E}_U e^{-\beta\ell(\sum_j \sqrt{s_j}g_jU_j-\epsilon)}} \xrightarrow{\beta\to\infty} \mathbb{E}\phi(v_*(0)). \tag{286}$$

The last limit uses the residual minimizer in (209) at $h = 0$. Together with (282), the saddle prediction is

$$\frac{1}{N}\sum_\mu \phi\Big(D^{-1/\alpha}\boldsymbol{x}_\mu^\top(\hat{\boldsymbol{w}}-\boldsymbol{w}^*)-\epsilon_\mu\Big) \xrightarrow{\text{p}} \mathbb{E}\phi(v_*(0)), \tag{287}$$

$$D^{-1/\alpha}\boldsymbol{x}_\mu^\top\hat{\boldsymbol{w}} - y_\mu = D^{-1/\alpha}\boldsymbol{x}_\mu^\top(\hat{\boldsymbol{w}}-\boldsymbol{w}^*)-\epsilon_\mu \xrightarrow{\text{d}} v_{\mu,*}(0). \tag{288}$$

This proves the training-residual law stated in (73). Combining (279) and (288), the observable content of the RS saddle is

$$\begin{aligned}\hat{w}_i - w_i^* &\xrightarrow{\text{d}} u_{i,*}(0),\\ D^{-1/\alpha}\boldsymbol{x}_\mu^\top\hat{\boldsymbol{w}} - y_\mu &\xrightarrow{\text{d}} v_{\mu,*}(0).\end{aligned} \tag{289}$$

For Huber loss, (203) shows

$$v_*(0) = \pm\delta \quad\Longrightarrow\quad p = \ell'(\pm\delta) = \pm\delta, \qquad \sqrt{Q}\,\xi - \epsilon = \pm\delta - \mathcal{R}'(\pm\delta). \tag{290}$$

Conditional on $(Q,\xi,\mathcal{R})$, the noise has a density, so

$$\Pr(v_*(0) = \pm\delta \mid Q,\xi,\mathcal{R}) \le \Pr\Big(\epsilon = \sqrt{Q}\,\xi \mp \delta + \mathcal{R}'(\pm\delta)\,\Big|\,Q,\xi,\mathcal{R}\Big) = 0. \tag{291}$$

Continuous approximations to $\phi(v) = \mathbf{1}_{\{|v|<\delta\}}$ in (287) therefore give

$$\frac{1}{N}\sum_\mu \mathbf{1}_{\{|D^{-1/\alpha}\boldsymbol{x}_\mu^\top\hat{\boldsymbol{w}}-y_\mu|<\delta\}} \xrightarrow{\text{p}} \Pr(|v_*(0)|<\delta). \tag{292}$$

D Derivation of Heavy-Tail Universality

In this section, we derive Result 5. The replica calculation uses the feature distribution when averaging one training example in Equation (83). We show that a general feature distribution satisfying Equation (18) gives the same average up to $o_D(1)$. The remaining saddle-point calculation is then exactly the one in Appendix C, which gives the university law.

Derivation of Result 5. Let the generative function of X as $\phi_X(t) = \mathbb{E}e^{itX}$. Symmetry and the normalized tail in Equation (18) give

$$\log\phi_X(t) = -|t|^\alpha + |t|^\alpha\,\eta(t), \tag{293}$$

where $\eta(t) \longrightarrow 0$ as $t\to 0$.

For the replicated empirical measures considered in the saddle calculation, assume for each fixed n and $\boldsymbol{k}\in\mathbb{R}^n$ that

$$D^{-1/\alpha}\max_{1\le i\le D}\left|\boldsymbol{k}^\top\boldsymbol{u}_i\right| \longrightarrow 0 \tag{294}$$

$$\sup_D \frac{1}{D}\sum_{i=1}^D \left|\boldsymbol{k}^\top\boldsymbol{u}_i\right|^\alpha < \infty. \tag{295}$$

Independence across feature coordinates and Equation (293) then give

$$\mathbb{E}\exp\left(iD^{-1/\alpha}\sum_{i=1}^{D}X_i\boldsymbol{k}^\top\boldsymbol{u}_i\right)=\exp\left(-\frac{1}{D}\sum_{i=1}^{D}\left|\boldsymbol{k}^\top\boldsymbol{u}_i\right|^\alpha+o_D(1)\right) \tag{296}$$

$$=\exp\left(-\int_{\mathbb{R}^n}\rho_D(\mathrm{d}\boldsymbol{v})\left|\boldsymbol{k}^\top\boldsymbol{v}\right|^\alpha+o_D(1)\right). \tag{297}$$

Indeed, the absolute remainder in the exponent is bounded by

$$\max_{1\le i\le D}\left|\eta\Big(D^{-1/\alpha}\boldsymbol{k}^\top\boldsymbol{u}_i\Big)\right|\frac{1}{D}\sum_{i=1}^{D}\left|\boldsymbol{k}^\top\boldsymbol{u}_i\right|^\alpha.$$

which vanishes by Equation (295). Equation (297) is Equation (83) up to $o_D(1)$. Assuming this error is uniform under the Fourier integral in Equation (88), the one-sample factor satisfies

$$\mathcal{I}_{n,D}^{P_X}[\rho_D]=\mathcal{I}_n^{S\alpha S(1)}[\rho_D]+o_D(1). \tag{298}$$

The other terms in the replica action do not depend on the feature distribution. Since $N/D\to\kappa$, Equation (298) therefore gives

$$\mathcal{S}_n^{P_X}[\rho,\hat\rho]=\mathcal{S}_n^{S\alpha S(1)}[\rho,\hat\rho]+o_D(1). \tag{299}$$

From Equation (299) onward, the saddle variation, replica-symmetric ansatz, replica limit, and zero-temperature limit are exactly those in Appendix C. The RS solution therefore gives Equation (19). □

E Derivation of the Data-Rich Scaling and Tail Reliability

In this section, we prove Results 6. Write $C:=C_\alpha^{\mathrm{S}}$ for the tail constant in Equation (18). The coefficient one-body equation and the sample-side saddle equations first determine the scale of its solution and hence the decay of $q_\alpha(\kappa)$. The stable prediction distribution then gives the probability of a large prediction error.

E.1 Coefficient scale from the one-body equation

For quantities in Appendix B, a subscript κ below indicates the fixed point at sample ratio κ, and $u_{\kappa,*}(h)$ denotes one copy of the coefficient solution (52) at that fixed point. At zero external field, this one-body problem is

$$u_{\kappa,*}(0)=\operatorname*{argmin}_{u\in\mathbb{R}}\left\{\lambda[\gamma(w^*+u)-\gamma(w^*)]+\sqrt{\hat Q_\kappa}\,\zeta_\kappa u+\hat{\mathcal{R}}_\kappa(u)\right\}. \tag{300}$$

Its optimality condition is

$$0\in\lambda\partial\gamma(w^*+u_{\kappa,*}(0))+\sqrt{\hat Q_\kappa}\,\zeta_\kappa+\hat{\mathcal{R}}'_\kappa(u_{\kappa,*}(0)). \tag{301}$$

Thus the scale of the coefficient displacement is determined by the stationarity equation involving the regularization derivative and the complete data derivative $\sqrt{\hat Q_\kappa}\,\zeta_\kappa+\hat{\mathcal{R}}'_\kappa(u)$.

First, we evaluate the scale of $\sqrt{\hat Q_\kappa}\,\zeta_\kappa+\hat{\mathcal{R}}'_\kappa(u)$. Equations (61) and (62) express the two parts of this data derivative through the same sample-side one-body solutions. Adding them cancels their values at zero and gives

$$\sqrt{\hat Q_\kappa}\,\zeta_\kappa+\hat{\mathcal{R}}'_\kappa(u)=\sum_{l\ge1}\sqrt{\hat s_{\kappa,l}}\,\hat g_l\ell'\Big(v_{\kappa,l,*}\Big(\sqrt{\hat s_{\kappa,l}}\,\hat g_l u\Big)\Big). \tag{302}$$

The Poisson points at ratio κ satisfy

$$\#\{l:\hat s_{\kappa,l}>t\}\overset{\mathrm{d}}{=}\mathrm{Poisson}\left(\frac{2\kappa d_\alpha}{\alpha}t^{-\alpha/2}\right). \tag{303}$$

If $\{\hat{s}_{1,l}\}_{l\geq 1}$ denotes the same process at ratio one, then

$$\#\Big\{l : \kappa^{2/\alpha}\hat{s}_{1,l} > t\Big\} = \#\Big\{l : \hat{s}_{1,l} > \kappa^{-2/\alpha}t\Big\} \tag{304}$$

$$\stackrel{\mathrm{d}}{=} \mathrm{Poisson}\left(\frac{2d_\alpha}{\alpha}\Big(\kappa^{-2/\alpha}t\Big)^{-\alpha/2}\right) \tag{305}$$

$$= \mathrm{Poisson}\left(\frac{2\kappa d_\alpha}{\alpha}t^{-\alpha/2}\right). \tag{306}$$

Hence the two point processes and their square roots obey

$$\{\hat{s}_{\kappa,l}\}_{l\geq 1} \stackrel{\mathrm{d}}{=} \Big\{\kappa^{2/\alpha}\hat{s}_{1,l}\Big\}_{l\geq 1}, \tag{307}$$

$$\Big\{\sqrt{\hat{s}_{\kappa,l}}\Big\}_{l\geq 1} \stackrel{\mathrm{d}}{=} \Big\{\kappa^{1/\alpha}\sqrt{\hat{s}_{1,l}}\Big\}_{l\geq 1}. \tag{308}$$

Both occurrences of $\sqrt{\hat{s}_{\kappa,l}}$ are normalized by evaluating the derivative at $\kappa^{-1/\alpha}u$ and dividing its value by $\kappa^{1/\alpha}$. Define this normalized derivative by

$$\Phi_\kappa(u) := \kappa^{-1/\alpha}\left[\sqrt{\hat{Q}_\kappa}\,\zeta_\kappa + \hat{\mathcal{R}}'_\kappa\Big(\kappa^{-1/\alpha}u\Big)\right] \tag{309}$$

$$\stackrel{\mathrm{d}}{=} \sum_{l\geq 1}\sqrt{\hat{s}_{1,l}}\,\hat{g}_l\ell'\Big(v_{\kappa,l,*}\Big(\sqrt{\hat{s}_{1,l}}\,\hat{g}_l u\Big)\Big). \tag{310}$$

The right side contains no explicit powers of κ. It depends on κ only through the sample-side solutions $v_{\kappa,l,*}$, whose distribution is determined by the fixed point. In the globally Lipschitz case, if L_ℓ is a Lipschitz constant of ℓ, then

$$\Big|\ell'\Big(v_{\kappa,l,*}\Big(\sqrt{\hat{s}_{1,l}}\,\hat{g}_l u\Big)\Big)\Big| \leq L_\ell. \tag{311}$$

In the squared-loss case, Equation (233) instead gives

$$\Big|v_{\kappa,l,*}\Big(\sqrt{\hat{s}_{1,l}}\,\hat{g}_l u\Big) - v_{\kappa,l,*}(0)\Big| \leq \sqrt{\hat{s}_{1,l}}\,|\hat{g}_l u|\,. \tag{312}$$

Under the coupling in Equation (308), the finite sample-side scale gives

$$\sum_{l\geq 1}\hat{s}_{1,l}v_{\kappa,l,*}(0)^2 = \kappa^{-2/\alpha}\hat{Q}_\kappa < \infty. \tag{313}$$

Together with $\sum_l \hat{s}_{1,l}\hat{g}_l^2 < \infty$ almost surely, Equation (312) gives the same local control of the Poisson series in Equation (310).

The data-rich limit of the RS fixed point gives local-uniform convergence of Φ_κ to a continuous random monotone function Φ with an almost surely unique zero $U \in \mathbb{R}$ and

$$0 < \mathbb{E}\,|U|^\alpha < \infty. \tag{314}$$

Derivation of Result 6. Substituting the definition of Φ_κ into Equation (301) and dividing by $\kappa^{1/\alpha}$ gives

$$0 \in \frac{\lambda}{\kappa^{1/\alpha}}\partial\gamma(w^* + u_{\kappa,*}(0)) + \Phi_\kappa\Big(\kappa^{1/\alpha}u_{\kappa,*}(0)\Big). \tag{315}$$

The finite convex regularizer has a locally bounded subdifferential. Thus, for every $M < \infty$,

$$\frac{\lambda}{\kappa^{1/\alpha}}\sup_{|u|\leq M}\sup_{s\in\partial\gamma(w^*+\kappa^{-1/\alpha}u)}|s| \longrightarrow 0. \tag{316}$$

The local-uniform RS limit of Φ_κ makes the data term in Equation (315) remain of order one. Each Φ_κ is nondecreasing because it is the rescaled subgradient of the convex data contribution; its local-uniform limit Φ is therefore also nondecreasing. Since Φ is continuous and monotone with unique zero U, for every $\varepsilon > 0$,

$$\Phi(U-\varepsilon) < 0 < \Phi(U+\varepsilon) \quad \text{almost surely.} \tag{317}$$

The locally uniform $o(1)$ regularizer term preserves these signs in Equation (315). Monotonicity therefore implies that its zero lies between $U - \varepsilon$ and $U + \varepsilon$ with probability tending to one. Hence

$$\kappa^{1/\alpha} u_{\kappa,*}(0) \xrightarrow{\mathrm{d}} U, \tag{318}$$

where U is the solution of $\Phi(U) = 0$. Taking the corresponding RS order-parameter limit gives convergence of the α-moments:

$$\kappa q_\alpha(\kappa) = \mathbb{E}\left|\kappa^{1/\alpha} u_{\kappa,*}(0)\right|^\alpha \longrightarrow \mathbb{E}\,|U|^\alpha =: c_\alpha. \tag{319}$$

Hence

$$q_\alpha(\kappa) = \frac{c_\alpha}{\kappa} + o\big(\kappa^{-1}\big). \tag{320}$$

Proposition 2 now gives

$$\kappa^{1/\alpha} q_\alpha(\kappa)^{1/\alpha} S_\alpha = \big(\kappa q_\alpha(\kappa)\big)^{1/\alpha} S_\alpha \xrightarrow{\mathrm{d}} c_\alpha^{1/\alpha} S_\alpha. \tag{321}$$

Since $|S_\alpha|$ has a finite positive median,

$$\operatorname{median}\Big(q_\alpha(\kappa)^{1/\alpha}\,|S_\alpha|\Big) = q_\alpha(\kappa)^{1/\alpha}\operatorname{median}(|S_\alpha|) = \Theta\Big(\kappa^{-1/\alpha}\Big). \tag{322}$$

The Gaussian fixed-point equations give $q_2^{\mathrm{G}}(\kappa) = c_2/\kappa + o(\kappa^{-1})$. Since $|S_2|$ also has a finite positive median,

$$\operatorname{median}\Big(q_2^{\mathrm{G}}(\kappa)^{1/2}\,|S_2|\Big) = \Theta\Big(\kappa^{-1/2}\Big). \tag{323}$$

This proves the median statements. □

Fixed-threshold part of Result 6. For every fixed $t > 0$, Equation (320) gives

$$t q_\alpha(\kappa)^{-1/\alpha} = t\left(\frac{\kappa}{c_\alpha}\right)^{1/\alpha}(1 + o(1)) \longrightarrow \infty. \tag{324}$$

The stable-tail relation $\Pr(|S_\alpha| > x) \sim C x^{-\alpha}$ can therefore be applied at this threshold:

$$\Pr\Big(q_\alpha(\kappa)^{1/\alpha}\,|S_\alpha| > t\Big) = \Pr\Big(|S_\alpha| > t q_\alpha(\kappa)^{-1/\alpha}\Big) \tag{325}$$

$$\sim C q_\alpha(\kappa) t^{-\alpha} \tag{326}$$

$$\sim \frac{C c_\alpha}{\kappa t^\alpha}. \tag{327}$$

This is Equation (21).

For Gaussian features, write $S_2 = \sqrt{2}Z$ with $Z \sim \mathcal{N}(0,1)$. The standard Gaussian tail asymptotic gives

$$\Pr\Big(q_2^{\mathrm{G}}(\kappa)^{1/2}\,|S_2| > t\Big) \sim \frac{2\sqrt{q_2^{\mathrm{G}}(\kappa)}}{t\sqrt{\pi}} \exp\left(-\frac{t^2}{4 q_2^{\mathrm{G}}(\kappa)}\right). \tag{328}$$

Therefore

$$\frac{1}{\kappa}\log\Pr\Big(q_2^{\mathrm{G}}(\kappa)^{1/2}\,|S_2| > t\Big) = -\frac{t^2}{4\kappa q_2^{\mathrm{G}}(\kappa)} + O\left(\frac{\log\kappa}{\kappa}\right) \longrightarrow -\frac{t^2}{4c_2}, \tag{329}$$

which proves the Gaussian statement in Result 6. □

F Supporting Calculations for the Bayes Ridge Coefficient Error

In this section, we prove Result 7. We specialize the two fixed-point updates in Appendix B to the matched quadratic model and then expand the resulting curvature equations as $\kappa \to \infty$.

F.1 Quadratic specialization of the fixed point

Equations (24)–(27) show that the Bayes estimator is the regularized ERM with squared loss, quadratic regularization, and $\lambda = \sigma^2/\tau^2$. We therefore specialize the two updates in Appendix B to

$$\ell(r) = \frac{r^2}{2}, \quad \gamma(w) = \frac{w^2}{2}, \quad \lambda = \frac{\sigma^2}{\tau^2}. \tag{330}$$

Both fixed-point updates preserve the family of centered quadratic functions. We may therefore write the Bayes fixed point as

$$\mathcal{R}(h) = \frac{\mathcal{R}''}{2}h^2, \quad \hat{\mathcal{R}}(h) = \frac{\hat{\mathcal{R}}''}{2}h^2, \quad \mathcal{R}'' \le 0, \quad \hat{\mathcal{R}}'' \ge 0. \tag{331}$$

Substituting these forms into the two one-body problems and saddle equations gives their complete quadratic specialization. On the coefficient side, Equation (52) becomes

$$u_{j,*}(h) = \operatorname*{argmin}_{u\in\mathbb{R}}\left\{\frac{\lambda + \hat{\mathcal{R}}''_j}{2}u^2 + \left(\lambda w_j^* + \sqrt{\hat{Q}_j}\,\zeta_j - h\right)u\right\} \tag{332}$$

$$= \frac{h - \lambda w_j^* - \sqrt{\hat{Q}_j}\,\zeta_j}{\lambda + \hat{\mathcal{R}}''_j}. \tag{333}$$

In particular,

$$u_{j,*}\big(-\sqrt{s_j}\,g_j h\big) - u_{j,*}(0) = -\frac{\sqrt{s_j}\,g_j h}{\lambda + \hat{\mathcal{R}}''_j}. \tag{334}$$

Equations (53)–(55) therefore reduce to

$$Q = \sum_{j\ge 1} s_j u_{j,*}(0)^2 \tag{335}$$

$$\sqrt{Q}\,\xi = \sum_{j\ge 1}\sqrt{s_j}\,g_j u_{j,*}(0) \tag{336}$$

$$\mathcal{R}'' = -\sum_{j\ge 1}\frac{s_j g_j^2}{\lambda + \hat{\mathcal{R}}''_j} \le 0. \tag{337}$$

On the sample side, for $\mathcal{R}''_l < 0$, $(-\mathcal{R}_l)^*(h) = -h^2/(2\mathcal{R}''_l)$. Hence Equation (59) becomes

$$v_{l,*}(h) = \operatorname*{argmin}_{v\in\mathbb{R}}\left\{\frac{v^2}{2} - \frac{\big(\sqrt{Q_l}\,\xi_l - \epsilon_l + h - v\big)^2}{2\mathcal{R}''_l}\right\} \tag{338}$$

$$= \frac{\sqrt{Q_l}\,\xi_l - \epsilon_l + h}{1 - \mathcal{R}''_l}. \tag{339}$$

This solution satisfies

$$v_{l,*}\Big(\sqrt{\hat{s}_l}\,\hat{g}_l h\Big) - v_{l,*}(0) = \frac{\sqrt{\hat{s}_l}\,\hat{g}_l h}{1 - \mathcal{R}''_l}. \tag{340}$$

Since $\ell'(v) = v$, Equations (60)–(62) reduce to

$$\hat{Q} = \sum_{l\ge 1}\hat{s}_l v_{l,*}(0)^2 \tag{341}$$

$$\sqrt{\hat{Q}}\,\zeta = \sum_{l\ge 1}\sqrt{\hat{s}_l}\,\hat{g}_l v_{l,*}(0) \tag{342}$$

$$\hat{\mathcal{R}}'' = \sum_{l\ge 1}\frac{\hat{s}_l \hat{g}_l^2}{1 - \mathcal{R}''_l} \ge 0. \tag{343}$$

The one-body solutions still contain the coefficient-side field $\lambda w^* + \sqrt{\hat{Q}}\,\zeta$ and the sample-side field $\sqrt{Q}\,\xi - \epsilon$. Next, we substitute the one-body solutions into the two linear-field updates. Equations (333) and (336) give

$$\sqrt{Q}\,\xi - \epsilon = -\epsilon - \sum_{j\geq 1} \frac{\sqrt{s_j}\,g_j}{\lambda + \hat{\mathcal{R}}''_j}\Big(\lambda w^*_j + \sqrt{\hat{Q}_j}\,\zeta_j\Big), \tag{344}$$

while Equation (339) and the sample-side linear-field update give

$$\lambda w^* + \sqrt{\hat{Q}}\,\zeta = \lambda w^* + \sum_{l\geq 1} \frac{\sqrt{\hat{s}_l}\,\hat{g}_l}{1 - \mathcal{R}''_l}\Big(\sqrt{Q_l}\,\xi_l - \epsilon_l\Big). \tag{345}$$

The sources w^* and ϵ are independent centered Gaussian variables, and the two updates are linear. Hence the two updates preserve the family of conditionally centered Gaussian fields. At the fixed point, their conditional variances satisfy

$$\mathrm{Var}\Big(\sqrt{Q}\,\xi - \epsilon \mid \{s_j, g_j, \hat{\mathcal{R}}''_j\}_{j\geq 1}\Big) = \sigma^2 + \sum_{j\geq 1} \frac{s_j g_j^2}{\Big(\lambda + \hat{\mathcal{R}}''_j\Big)^2}\,\mathrm{Var}\Big(\lambda w^*_j + \sqrt{\hat{Q}_j}\,\zeta_j \,\Big|\, \hat{\mathcal{R}}''_j\Big) \tag{346}$$

$$\mathrm{Var}\Big(\lambda w^* + \sqrt{\hat{Q}}\,\zeta \mid \{\hat{s}_l, \hat{g}_l, \mathcal{R}''_l\}_{l\geq 1}\Big) = \lambda^2\tau^2 + \sum_{l\geq 1} \frac{\hat{s}_l \hat{g}_l^2}{(1 - \mathcal{R}''_l)^2}\,\mathrm{Var}\Big(\sqrt{Q_l}\,\xi_l - \epsilon_l \,\Big|\, \mathcal{R}''_l\Big). \tag{347}$$

The two curvature equations determine these variance recursions without additional quantities. Substituting the sample-side variance $\sigma^2(1 - \mathcal{R}'')$ or the coefficient-side variance $\sigma^2(\lambda + \hat{\mathcal{R}}'')$ cancels one power of the corresponding denominator, and the remaining sum equals the corresponding curvature. Explicitly,

$$\sigma^2 + \sum_{j\geq 1} \frac{s_j g_j^2}{\Big(\lambda + \hat{\mathcal{R}}''_j\Big)^2}\sigma^2\Big(\lambda + \hat{\mathcal{R}}''_j\Big) = \sigma^2\left(1 + \sum_{j\geq 1} \frac{s_j g_j^2}{\lambda + \hat{\mathcal{R}}''_j}\right) \tag{348}$$

$$= \sigma^2(1 - \mathcal{R}''), \tag{349}$$

$$\lambda^2\tau^2 + \sum_{l\geq 1} \frac{\hat{s}_l \hat{g}_l^2}{(1 - \mathcal{R}''_l)^2}\sigma^2(1 - \mathcal{R}''_l) = \sigma^2\lambda + \sigma^2 \sum_{l\geq 1} \frac{\hat{s}_l \hat{g}_l^2}{1 - \mathcal{R}''_l} \tag{350}$$

$$= \sigma^2\Big(\lambda + \hat{\mathcal{R}}''\Big), \tag{351}$$

where we used Equations (337) and (343), together with $\lambda^2\tau^2 = \sigma^2\lambda$, which follows from $\lambda = \sigma^2/\tau^2$. Thus the conditional distributions reduce to

$$\sqrt{Q}\,\xi - \epsilon \mid \mathcal{R}'' \sim \mathcal{N}\big(0, \sigma^2(1 - \mathcal{R}'')\big), \tag{352}$$

$$\lambda w^* + \sqrt{\hat{Q}}\,\zeta \mid \hat{\mathcal{R}}'' \sim \mathcal{N}\Big(0, \sigma^2\Big(\lambda + \hat{\mathcal{R}}''\Big)\Big). \tag{353}$$

Substituting the second relation into Equation (333) gives

$$u_*(0) \mid \hat{\mathcal{R}}'' \sim \mathcal{N}\left(0, \frac{\sigma^2}{\lambda + \hat{\mathcal{R}}''}\right). \tag{354}$$

The coefficient distribution (69), Equation (354), and Equation (28) now reduce the coefficient error to a negative moment of $\lambda + \hat{\mathcal{R}}''$. Indeed,

$$\mathbb{E}\Big[\,|u_*(0)|^\alpha \,\Big|\, \hat{\mathcal{R}}''\Big] = \frac{2^{\alpha/2}\Gamma((\alpha+1)/2)}{\sqrt{\pi}}\sigma^\alpha\Big(\lambda + \hat{\mathcal{R}}''\Big)^{-\alpha/2}. \tag{355}$$

Averaging over $\hat{\mathcal{R}}''$ gives

$$q_{\alpha,\mathrm{B}}(\kappa) = \frac{2^{\alpha/2}\Gamma((\alpha+1)/2)}{\sqrt{\pi}}\sigma^\alpha\mathbb{E}\Big(\lambda + \hat{\mathcal{R}}''\Big)^{-\alpha/2}. \tag{356}$$

F.2 DATA-RICH EXPANSION

It remains to evaluate the negative curvature moment in Equation (356). The marked-Poisson representation in Equation (121) implies that the signed points $\sqrt{s_j}g_j$ have total tail constant C_α^{S}. Define the squared points

$$Y_j := s_j g_j^2, \quad \hat{Y}_l := \hat{s}_l \hat{g}_l^2. \tag{357}$$

The mapping theorem shows that $\{Y_j\}_{j\geq 1}$ and $\{\hat{Y}_l\}_{l\geq 1}$ are positive Poisson processes with respective Lévy measures

$$\nu(\mathrm{d}y) = \frac{\alpha}{2} C_\alpha^{\mathrm{S}} y^{-1-\alpha/2}\,\mathrm{d}y, \quad \hat{\nu}(\mathrm{d}y) = \kappa\nu(\mathrm{d}y). \tag{358}$$

Equation (343) and the exponential formula for a marked Poisson process first give

$$\mathbb{E}\exp\Big(-t\hat{\mathcal{R}}''\Big) = \mathbb{E}\exp\left(-t\sum_{l\geq 1}\frac{\hat{Y}_l}{1-\mathcal{R}_l''}\right) \tag{359}$$

$$= \exp\left\{-\kappa\frac{\alpha}{2}C_\alpha^{\mathrm{S}}\int_0^\infty \mathbb{E}\left[1-\exp\left(-\frac{ty}{1-\mathcal{R}''}\right)\right] y^{-1-\alpha/2}\,\mathrm{d}y\right\}. \tag{360}$$

The integral in this exponent comes from the Lévy measure. For fixed $r \leq 0$ and $t > 0$, the change of variables $u = ty/(1-r)$ and integration by parts yield

$$\frac{\alpha}{2}\int_0^\infty \Big(1-e^{-ty/(1-r)}\Big) y^{-1-\alpha/2}\,\mathrm{d}y = \left(\frac{t}{1-r}\right)^{\alpha/2}\frac{\alpha}{2}\int_0^\infty \big(1-e^{-u}\big)u^{-1-\alpha/2}\,\mathrm{d}u \tag{361}$$

$$= \left(\frac{t}{1-r}\right)^{\alpha/2}\int_0^\infty e^{-u}u^{-\alpha/2}\,\mathrm{d}u \tag{362}$$

$$= \Gamma(1-\alpha/2)\left(\frac{t}{1-r}\right)^{\alpha/2}. \tag{363}$$

Substituting this identity at $r = \mathcal{R}''$ gives

$$\mathbb{E}\exp\Big(-t\hat{\mathcal{R}}''\Big) = \exp\left\{-\kappa C_\alpha^{\mathrm{S}}\Gamma(1-\alpha/2)\mathbb{E}\left(\frac{t}{1-\mathcal{R}''}\right)^{\alpha/2}\right\} \tag{364}$$

$$= \exp\left\{-\kappa C_\alpha^{\mathrm{S}}\Gamma(1-\alpha/2)\mathbb{E}(1-\mathcal{R}'')^{-\alpha/2}t^{\alpha/2}\right\}, \tag{365}$$

Similarly, Equation (337) gives

$$\mathbb{E}\exp(t\mathcal{R}'') = \mathbb{E}\exp\left(-t\sum_{j\geq 1}\frac{Y_j}{\lambda+\hat{\mathcal{R}}_j''}\right) \tag{366}$$

$$= \exp\left\{-\frac{\alpha}{2}C_\alpha^{\mathrm{S}}\int_0^\infty \mathbb{E}\left[1-\exp\left(-\frac{ty}{\lambda+\hat{\mathcal{R}}''}\right)\right] y^{-1-\alpha/2}\,\mathrm{d}y\right\} \tag{367}$$

$$= \exp\left\{-C_\alpha^{\mathrm{S}}\Gamma(1-\alpha/2)\mathbb{E}\Big(\lambda+\hat{\mathcal{R}}''\Big)^{-\alpha/2}t^{\alpha/2}\right\}. \tag{368}$$

We then substitute these Laplace transforms into

$$z^{-\alpha/2} = \frac{1}{\Gamma(\alpha/2)}\int_0^\infty t^{\alpha/2-1}e^{-tz}\,\mathrm{d}t, \qquad z > 0. \tag{369}$$

This gives the two closed moment equations

$$\mathbb{E}\Big(\lambda+\hat{\mathcal{R}}''\Big)^{-\alpha/2}=\frac{1}{\Gamma(\alpha/2)}\int_0^\infty t^{\alpha/2-1}\exp\left\{-\lambda t-\kappa C_\alpha^{\mathrm{S}}\Gamma(1-\alpha/2)\mathbb{E}(1-\mathcal{R}'')^{-\alpha/2}t^{\alpha/2}\right\}\mathrm{d}t, \tag{370}$$

$$\mathbb{E}(1-\mathcal{R}'')^{-\alpha/2}=\frac{1}{\Gamma(\alpha/2)}\int_0^\infty t^{\alpha/2-1}\exp\left\{-t-C_\alpha^{\mathrm{S}}\Gamma(1-\alpha/2)\mathbb{E}\Big(\lambda+\hat{\mathcal{R}}''\Big)^{-\alpha/2}t^{\alpha/2}\right\}\mathrm{d}t. \tag{371}$$

Using $\lambda=\sigma^2/\tau^2$, making the changes of variables $u=t^{\alpha/2}$ and $v=t^{\alpha/2}$ in the two respective integrals, substituting Equation (371) into Equation (370), and using Equation (356) gives Equation (29).

Since $\hat{\mathcal{R}}''\geq 0$,

$$\mathbb{E}\Big(\lambda+\hat{\mathcal{R}}''\Big)^{-\alpha/2}\leq\lambda^{-\alpha/2}. \tag{372}$$

Equation (371) therefore gives the κ-independent lower bound

$$\mathbb{E}(1-\mathcal{R}'')^{-\alpha/2}\geq\frac{1}{\Gamma(\alpha/2)}\int_0^\infty t^{\alpha/2-1}\exp\left\{-t-C_\alpha^{\mathrm{S}}\Gamma(1-\alpha/2)\lambda^{-\alpha/2}t^{\alpha/2}\right\}\mathrm{d}t=:b_\lambda>0. \tag{373}$$

Substituting this lower bound into Equation (370) yields

$$\mathbb{E}\Big(\lambda+\hat{\mathcal{R}}''\Big)^{-\alpha/2}\leq\frac{1}{\Gamma(\alpha/2)}\int_0^\infty t^{\alpha/2-1}\exp\left\{-\kappa C_\alpha^{\mathrm{S}}\Gamma(1-\alpha/2)b_\lambda t^{\alpha/2}\right\}\mathrm{d}t \tag{374}$$

$$=\frac{1}{\kappa C_\alpha^{\mathrm{S}}\Gamma(1-\alpha/2)\Gamma(1+\alpha/2)b_\lambda}\longrightarrow 0. \tag{375}$$

Dominated convergence in Equation (371) then gives

$$\mathbb{E}(1-\mathcal{R}'')^{-\alpha/2}\longrightarrow 1. \tag{376}$$

Substituting this limit into Equation (370) and setting $t=\kappa^{-2/\alpha}s$ gives

$$\mathbb{E}\Big(\lambda+\hat{\mathcal{R}}''\Big)^{-\alpha/2}=\frac{1}{\kappa\Gamma(\alpha/2)}\int_0^\infty s^{\alpha/2-1}\times\exp\left\{-\lambda\kappa^{-2/\alpha}s-C_\alpha^{\mathrm{S}}\Gamma(1-\alpha/2)\mathbb{E}(1-\mathcal{R}'')^{-\alpha/2}s^{\alpha/2}\right\}\mathrm{d}s \tag{377}$$

$$=\frac{1}{\kappa\Gamma(\alpha/2)}\int_0^\infty s^{\alpha/2-1}\exp\left\{-C_\alpha^{\mathrm{S}}\Gamma(1-\alpha/2)s^{\alpha/2}\right\}\mathrm{d}s+o\big(\kappa^{-1}\big) \tag{378}$$

$$=\frac{1}{\kappa C_\alpha^{\mathrm{S}}\Gamma(1-\alpha/2)\Gamma(1+\alpha/2)}+o\big(\kappa^{-1}\big). \tag{379}$$

Substituting Equation (379) into Equation (356) gives

$$q_{\alpha,\mathrm{B}}(\kappa)=\frac{2^{\alpha/2}\Gamma((\alpha+1)/2)}{\sqrt{\pi}}\sigma^\alpha\left[\frac{1}{\kappa C_\alpha^{\mathrm{S}}\Gamma(1-\alpha/2)\Gamma(1+\alpha/2)}+o\big(\kappa^{-1}\big)\right] \tag{380}$$

$$=\frac{2^{\alpha/2}\Gamma((\alpha+1)/2)}{\alpha\sqrt{\pi}\,\Gamma(\alpha)}\frac{\sigma^\alpha}{\kappa}+o\big(\kappa^{-1}\big) \tag{381}$$

$$=\frac{2^{-\alpha/2}}{\Gamma(1+\alpha/2)}\frac{\sigma^\alpha}{\kappa}+o\big(\kappa^{-1}\big). \tag{382}$$

The last two equalities use the definition of C_α^{S} and the Gamma duplication formula.

F.3 Elliptical Bayes fixed point

Since $x_{\mu i} \sim S\alpha S(1)$, $\Pr(x_{\mu i}^2 > y) \sim C_\alpha^{\mathrm{S}} y^{-\alpha/2}$. The normalization in Equation (32) therefore gives R_μ^2 the positive $\alpha/2$-stable distribution

$$\begin{aligned}\mathbb{E}e^{-tR_\mu^2} &= \lim_{D\to\infty}\left[\mathbb{E}e^{-tD^{-2/\alpha}x_{\mu 1}^2}\right]^D \\ &= \exp\Big(-C_\alpha^{\mathrm{S}}\Gamma(1-\alpha/2)t^{\alpha/2}\Big) = \exp\left(-\frac{2^\alpha\Gamma((\alpha+1)/2)}{\sqrt{\pi}}t^{\alpha/2}\right).\end{aligned}$$

The stable-limit exponent follows from the squared-coordinate tail; its simplification uses the Gamma reflection and duplication formulas.

Let X^{ell} have rows $(\boldsymbol{x}_\mu^{\mathrm{ell}})^\top$. Write χ for the proportional-limit normalized resolvent

$$\chi = \lim_{D\to\infty}\frac{1}{D}\operatorname{Tr}\left(D^{-2/\alpha}(X^{\mathrm{ell}})^\top X^{\mathrm{ell}} + \frac{\sigma^2}{\tau^2}I_D\right)^{-1}. \tag{383}$$

In our notation, the Gaussian-prior/noise Bayes equations of Adomaityte et al. (2024) give the fixed point

$$1 = \frac{\sigma^2}{\tau^2}\chi + \kappa\left[1 - \mathbb{E}\frac{1}{1+\chi R_\mu^2}\right]. \tag{384}$$

Rotational invariance makes the diagonal posterior variances converge to $\sigma^2\chi$. Thus, for $Z \sim \mathcal{N}(0,1)$,

$$q_{\alpha,\mathrm{B}}^{\mathrm{ell}} = \mathbb{E}|Z|^\alpha\sigma^\alpha\chi^{\alpha/2} = \frac{2^{\alpha/2}\Gamma((\alpha+1)/2)}{\sqrt{\pi}}\sigma^\alpha\chi^{\alpha/2}. \tag{385}$$

Consequently, the prior term becomes

$$\frac{\sigma^2}{\tau^2}\chi = \frac{1}{\tau^2}\left[\frac{\sqrt{\pi}\,q_{\alpha,\mathrm{B}}^{\mathrm{ell}}}{2^{\alpha/2}\Gamma((\alpha+1)/2)}\right]^{2/\alpha}, \tag{386}$$

while the Laplace representation of $(1+\chi R_\mu^2)^{-1}$ gives

$$\mathbb{E}\frac{1}{1+\chi R_\mu^2} = \int_0^\infty \exp\left(-s - \frac{2^\alpha\Gamma((\alpha+1)/2)}{\sqrt{\pi}}\chi^{\alpha/2}s^{\alpha/2}\right)\mathrm{d}s \tag{387}$$

$$= \sigma^2\int_0^\infty \exp\Big(-\sigma^2 u - 2^{\alpha/2}q_{\alpha,\mathrm{B}}^{\mathrm{ell}}u^{\alpha/2}\Big)\,\mathrm{d}u. \tag{388}$$

Substituting these two expressions into the known fixed point yields Equation (33):

$$1 = \frac{1}{\tau^2}\left[\frac{\sqrt{\pi}\,q_{\alpha,\mathrm{B}}^{\mathrm{ell}}(\kappa)}{2^{\alpha/2}\Gamma((\alpha+1)/2)}\right]^{2/\alpha} + \kappa\left[1 - \sigma^2\int_0^\infty \exp\Big(-\sigma^2 u - 2^{\alpha/2}q_{\alpha,\mathrm{B}}^{\mathrm{ell}}(\kappa)u^{\alpha/2}\Big)\,\mathrm{d}u\right], \tag{389}$$

F.4 Comparison with elliptical features

Finally, we derive Result 8. Equation (389) gives the elliptical coefficient error, and the coordinate-wise independent coefficient error satisfies

$$\begin{aligned}q_{\alpha,\mathrm{B}}(\kappa) = \frac{2^{\alpha/2}\Gamma((\alpha+1)/2)\sigma^\alpha}{\sqrt{\pi}\,\Gamma(1+\alpha/2)}&\int_0^\infty \exp\left(-\frac{\sigma^2}{\tau^2}u^{2/\alpha}\right) \\ &\times\exp\left\{-\frac{\kappa 2^\alpha\Gamma((\alpha+1)/2)}{\sqrt{\pi}\,\Gamma(1+\alpha/2)}u\int_0^\infty\exp\left(-v^{2/\alpha} - \frac{2^{\alpha/2}q_{\alpha,\mathrm{B}}(\kappa)}{\sigma^\alpha}v\right)\mathrm{d}v\right\}\mathrm{d}u.\end{aligned} \tag{390}$$

Our strategy is to show that the right-hand side of Equation (389) is strictly increasing in the coefficient error q. We then use Equation (390) to show that this right-hand side, evaluated at $q = q_{\alpha,\mathrm{B}}(\kappa)$,

is strictly less than 1. Since it equals 1 at $q = q^{\mathrm{ell}}_{\alpha,\mathrm{B}}(\kappa)$, the desired ordering $q_{\alpha,\mathrm{B}}(\kappa) < q^{\mathrm{ell}}_{\alpha,\mathrm{B}}(\kappa)$ follows.

First, we rewrite the right-hand side of Equation (389) and show that it is strictly increasing in q. For any $q > 0$, the second term in Equation (389) is

$$\kappa \left[1 - \sigma^2 \int_0^\infty \exp\Big(-\sigma^2 u - 2^{\alpha/2} q u^{\alpha/2}\Big)\,\mathrm{d}u\right] \tag{391}$$

$$= \kappa \left[1 - \int_0^\infty \exp\left(-x - \frac{2^{\alpha/2} q}{\sigma^\alpha} x^{\alpha/2}\right)\mathrm{d}x\right] \tag{392}$$

$$= -\kappa \int_0^\infty e^{-x} \frac{\mathrm{d}}{\mathrm{d}x} \exp\left(-\frac{2^{\alpha/2} q}{\sigma^\alpha} x^{\alpha/2}\right)\mathrm{d}x \tag{393}$$

$$= \frac{\alpha 2^{\alpha/2} \kappa q}{2\sigma^\alpha} \int_0^\infty x^{\alpha/2-1} \exp\left(-x - \frac{2^{\alpha/2} q}{\sigma^\alpha} x^{\alpha/2}\right)\mathrm{d}x \tag{394}$$

$$= \frac{2^{\alpha/2} \kappa q}{\sigma^\alpha} \int_0^\infty \exp\left(-v^{2/\alpha} - \frac{2^{\alpha/2} q}{\sigma^\alpha} v\right)\mathrm{d}v. \tag{395}$$

Consequently, Equation (389) is equivalently

$$1 = \frac{1}{\tau^2}\left[\frac{\sqrt{\pi}\, q^{\mathrm{ell}}_{\alpha,\mathrm{B}}}{2^{\alpha/2}\Gamma((\alpha+1)/2)}\right]^{2/\alpha} + \frac{2^{\alpha/2}\kappa q^{\mathrm{ell}}_{\alpha,\mathrm{B}}}{\sigma^\alpha} \int_0^\infty \exp\left(-v^{2/\alpha} - \frac{2^{\alpha/2} q^{\mathrm{ell}}_{\alpha,\mathrm{B}}}{\sigma^\alpha} v\right)\mathrm{d}v. \tag{396}$$

The first term on the right-hand side of Equation (396) is strictly increasing in the coefficient error, and the second term satisfies

$$\frac{\mathrm{d}}{\mathrm{d}q}\kappa \left[1 - \int_0^\infty \exp\left(-x - \frac{2^{\alpha/2} q}{\sigma^\alpha} x^{\alpha/2}\right)\mathrm{d}x\right] \tag{397}$$

$$= \frac{2^{\alpha/2}\kappa}{\sigma^\alpha} \int_0^\infty x^{\alpha/2} \exp\left(-x - \frac{2^{\alpha/2} q}{\sigma^\alpha} x^{\alpha/2}\right)\mathrm{d}x > 0. \tag{398}$$

Thus the entire right-hand side is strictly increasing in q.

We next put Equation (390) in the same form. For $q_{\alpha,\mathrm{B}} = q_{\alpha,\mathrm{B}}(\kappa)$, let T be the positive $\alpha/2$-stable random variable with

$$\mathbb{E}e^{-xT} = \exp\left\{-\frac{\kappa 2^\alpha \Gamma((\alpha+1)/2)}{\sqrt{\pi}\,\Gamma(1+\alpha/2)} x^{\alpha/2} \int_0^\infty \exp\left(-v^{2/\alpha} - \frac{2^{\alpha/2} q_{\alpha,\mathrm{B}}}{\sigma^\alpha} v\right)\mathrm{d}v\right\}. \tag{399}$$

Using this notation gives

$$\frac{\sqrt{\pi}\, q_{\alpha,\mathrm{B}}}{2^{\alpha/2}\Gamma((\alpha+1)/2)\sigma^\alpha} \tag{400}$$

$$= \frac{1}{\Gamma(1+\alpha/2)} \int_0^\infty \exp\left(-\frac{\sigma^2}{\tau^2} u^{2/\alpha}\right)\mathbb{E}e^{-u^{2/\alpha}T}\,\mathrm{d}u \tag{401}$$

$$= \frac{1}{\Gamma(\alpha/2)} \int_0^\infty x^{\alpha/2-1} \exp\left(-\frac{\sigma^2}{\tau^2} x\right)\mathbb{E}e^{-xT}\,\mathrm{d}x \tag{402}$$

$$= \mathbb{E}\left(\frac{\sigma^2}{\tau^2} + T\right)^{-\alpha/2}. \tag{403}$$

From the same Laplace transform, we have the following identity:

$$\mathbb{E}\frac{T}{\sigma^2/\tau^2+T} \tag{404}$$

$$=\int_0^\infty \exp\left(-\frac{\sigma^2}{\tau^2}x\right)\mathbb{E}\big(Te^{-xT}\big)\,\mathrm{d}x \tag{405}$$

$$=-\int_0^\infty \exp\left(-\frac{\sigma^2}{\tau^2}x\right)\frac{\mathrm{d}}{\mathrm{d}x}\mathbb{E}e^{-xT}\,\mathrm{d}x \tag{406}$$

$$=\frac{\alpha\kappa 2^\alpha\Gamma((\alpha+1)/2)}{2\sqrt{\pi}\,\Gamma(1+\alpha/2)}\int_0^\infty \exp\left(-v^{2/\alpha}-\frac{2^{\alpha/2}q_{\alpha,\mathrm{B}}}{\sigma^\alpha}v\right)\mathrm{d}v\int_0^\infty x^{\alpha/2-1}\exp\left(-\frac{\sigma^2}{\tau^2}x\right)\mathbb{E}e^{-xT}\,\mathrm{d}x \tag{407}$$

$$=\frac{\kappa 2^\alpha\Gamma((\alpha+1)/2)}{\sqrt{\pi}\,\Gamma(1+\alpha/2)}\int_0^\infty \exp\left(-v^{2/\alpha}-\frac{2^{\alpha/2}q_{\alpha,\mathrm{B}}}{\sigma^\alpha}v\right)\mathrm{d}v\int_0^\infty \exp\left(-\frac{\sigma^2}{\tau^2}u^{2/\alpha}\right)\mathbb{E}e^{-u^{2/\alpha}T}\,\mathrm{d}u \tag{408}$$

$$=\frac{2^{\alpha/2}\kappa q_{\alpha,\mathrm{B}}}{\sigma^\alpha}\int_0^\infty \exp\left(-v^{2/\alpha}-\frac{2^{\alpha/2}q_{\alpha,\mathrm{B}}}{\sigma^\alpha}v\right)\mathrm{d}v. \tag{409}$$

The last equality uses Equation (390). Therefore

$$1=\mathbb{E}\frac{\sigma^2/\tau^2+T}{\sigma^2/\tau^2+T} \tag{410}$$

$$=\frac{\sigma^2}{\tau^2}\mathbb{E}\left(\frac{\sigma^2}{\tau^2}+T\right)^{-1}+\mathbb{E}\frac{T}{\sigma^2/\tau^2+T} \tag{411}$$

$$=\frac{\sigma^2}{\tau^2}\mathbb{E}\left(\frac{\sigma^2}{\tau^2}+T\right)^{-1}+\frac{2^{\alpha/2}\kappa q_{\alpha,\mathrm{B}}}{\sigma^\alpha}\int_0^\infty \exp\left(-v^{2/\alpha}-\frac{2^{\alpha/2}q_{\alpha,\mathrm{B}}}{\sigma^\alpha}v\right)\mathrm{d}v. \tag{412}$$

We now use Equation (412) and the strict monotonicity of L^p norms for $0<\alpha/2<1$:

$$1=\frac{\sigma^2}{\tau^2}\mathbb{E}\left(\frac{\sigma^2}{\tau^2}+T\right)^{-1}+\frac{2^{\alpha/2}\kappa q_{\alpha,\mathrm{B}}}{\sigma^\alpha}\int_0^\infty \exp\left(-v^{2/\alpha}-\frac{2^{\alpha/2}q_{\alpha,\mathrm{B}}}{\sigma^\alpha}v\right)\mathrm{d}v \tag{413}$$

$$>\frac{\sigma^2}{\tau^2}\left[\mathbb{E}\left(\frac{\sigma^2}{\tau^2}+T\right)^{-\alpha/2}\right]^{2/\alpha}+\frac{2^{\alpha/2}\kappa q_{\alpha,\mathrm{B}}}{\sigma^\alpha}\int_0^\infty \exp\left(-v^{2/\alpha}-\frac{2^{\alpha/2}q_{\alpha,\mathrm{B}}}{\sigma^\alpha}v\right)\mathrm{d}v \tag{414}$$

$$=\frac{1}{\tau^2}\left[\frac{\sqrt{\pi}\,q_{\alpha,\mathrm{B}}}{2^{\alpha/2}\Gamma((\alpha+1)/2)}\right]^{2/\alpha}+\frac{2^{\alpha/2}\kappa q_{\alpha,\mathrm{B}}}{\sigma^\alpha}\int_0^\infty \exp\left(-v^{2/\alpha}-\frac{2^{\alpha/2}q_{\alpha,\mathrm{B}}}{\sigma^\alpha}v\right)\mathrm{d}v. \tag{415}$$

The strict inequality follows from the strict monotonicity of L^p norms for $0<\alpha/2<1$. It is strict because the Laplace exponent in Equation (399) is positive for $\kappa>0$, so T is nonconstant. The last equality uses Equation (403).

Thus the elliptical right-hand side at $q_{\alpha,\mathrm{B}}$ is strictly less than one. By the strict monotonicity established above, it equals one only at a larger coefficient error:

$$q^{\mathrm{ell}}_{\alpha,\mathrm{B}}(\kappa)>q_{\alpha,\mathrm{B}}(\kappa), \tag{416}$$

which proves Equation (34).

It remains to compare the leading data-rich terms of $q^{\mathrm{ell}}_{\alpha,\mathrm{B}}(\kappa)$. Equation (396) first implies $q^{\mathrm{ell}}_{\alpha,\mathrm{B}}(\kappa)\to 0$. Hence

$$\int_0^\infty \exp\left(-v^{2/\alpha}-\frac{2^{\alpha/2}q^{\mathrm{ell}}_{\alpha,\mathrm{B}}(\kappa)}{\sigma^\alpha}v\right)\mathrm{d}v=\int_0^\infty e^{-v^{2/\alpha}}\,\mathrm{d}v+o(1) \tag{417}$$

$$=\frac{\alpha}{2}\int_0^\infty e^{-x}x^{\alpha/2-1}\,\mathrm{d}x+o(1) \tag{418}$$

$$=\Gamma(1+\alpha/2)+o(1). \tag{419}$$

Substitution into the same fixed point gives

$$1 = \frac{1}{\tau^2}\left[\frac{\sqrt{\pi}\, q^{\mathrm{ell}}_{\alpha,\mathrm{B}}(\kappa)}{2^{\alpha/2}\Gamma((\alpha+1)/2)}\right]^{2/\alpha} + \frac{2^{\alpha/2}\kappa q^{\mathrm{ell}}_{\alpha,\mathrm{B}}(\kappa)}{\sigma^\alpha}\left[\Gamma(1+\alpha/2) + o(1)\right] \tag{420}$$

$$= o(1) + \frac{2^{\alpha/2}\kappa q^{\mathrm{ell}}_{\alpha,\mathrm{B}}(\kappa)}{\sigma^\alpha}\left[\Gamma(1+\alpha/2) + o(1)\right], \tag{421}$$

which gives

$$q^{\mathrm{ell}}_{\alpha,\mathrm{B}}(\kappa) = \frac{2^{-\alpha/2}}{\Gamma(1+\alpha/2)}\frac{\sigma^\alpha}{\kappa} + o\big(\kappa^{-1}\big). \tag{422}$$

Together with Equation (30), this proves Equation (35).