# Universality of Singular Complexity for Hyvärinen Generalized Bayes: Exact Transfer in Gaussian Factor Analysis

**Manoj Saravanan**[1] **Rohit Kumar Salla**[1]

[1]Virginia Tech, Department of Electrical and Computer Engineering, Blacksburg, VA, USA

## Abstract

Singular statistical models are governed asymptotically by local birational invariants such as the real log canonical threshold (RLCT), not by ambient parameter dimension. We study this complexity under generalized Bayes updates based on the Hyvärinen loss. Our first result is a local comparison theorem: nonnegative analytic excess-loss germs that are locally comparable near a common zero set have the same local RLCT pair. As a corollary, losses with the same analytic minimizer map and a nondegenerate quadratic germ belong to the same singular universality class. We then show that, in zero-mean Gaussian covariance models, the population excess Gaussian log-loss and the excess Hyvärinen loss are both locally equivalent to $\|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2$. Hence ordinary Gaussian Bayes and Hyvärinen generalized Bayes have identical local RLCT pairs at every covariance fiber. Applying this to Gaussian factor analysis yields exact transfer of all currently known local learning-coefficient results. In the one-factor model, the Hyvärinen posterior therefore has the explicit coefficients $p$, $(2p-1)/2$, and $3p/4$ on the corresponding covariance strata. Numerical experiments confirm the local quadratic equivalence and the cancellation of the shared $\lambda \log n$ term in paired free-energy differences.

## 1. Introduction

Singular statistical models lie outside classical Laplace asymptotics. In latent-variable models, mixtures, hierarchical models with nonidentifiability, and related analytically singular families, the leading complexity is governed by the real log canonical threshold (RLCT), or learning coefficient, rather than by ambient parameter dimension (Watanabe, 2009, 2013). Gaussian factor analysis is a canonical example: the covariance map $(\psi, \Lambda) \mapsto \mathrm{diag}(\psi) + \Lambda\Lambda^\top$ is highly noninjective, and different covariance fibers carry different learning coefficients. Recent work gives the first systematic singular-learning-theoretic analysis of this model, including exact formulas in important cases and general upper bounds (Drton et al., 2025).

A separate line of work studies *generalized Bayes*, where the negative log-likelihood is replaced by a loss and the posterior is updated by exponential tilting (Bissiri et al., 2016). A particularly important example is the Hyvärinen score, whose score-matching form depends only on derivatives of the log density and therefore avoids normalizing constants (Hyvärinen, 2005); this makes it attractive when likelihood normalizations are unavailable or inconvenient (Jewson and Rossell, 2022). At the same time, outside the log-loss case such updates are generally decision posteriors rather than ordinary Bayes posteriors (McAlinn and Takanashi, 2026). This raises a precise local question: if the Gaussian log-loss is replaced by the Hyvärinen loss in a singular model, does the RLCT change?

There is no a priori reason that it should not. The RLCT is determined by the vanishing structure of the population excess-loss germ near its minimizer fiber, and a different loss can in principle sharpen or flatten that germ. Indeed, our sharpness result below shows that when one loss is locally comparable to a higher power of another, the learning coefficient rescales by the reciprocal power. Any universality statement therefore has to be proved at the level of local excess-loss geometry, rather than inferred from superficial similarity between update rules.

Our answer is no for analytic Gaussian covariance models. First, we prove a local comparison theorem: nonnegative analytic excess-loss germs that are locally comparable near a common zero set have the same local RLCT pair. As a corollary, any two losses with the same analytic minimizer map and a positive-definite quadratic germ lie in the same singular universality class. Second, we show that in zero-mean Gaussian covariance models the Gaussian excess log-loss and the excess Hyvärinen loss are both locally equivalent to $\|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2$. Hence ordinary Gaussian Bayes and Hyvärinen generalized Bayes have identical local RLCT pairs at every covariance fiber. Third, applying this transfer principle to factor analysis yields exact inheritance of all currently known Bayes-side RLCT results. In particular, in the one-factor model the Hyvärinen posterior has learning coefficients

$$p, \qquad \frac{2p-1}{2}, \qquad \frac{3p}{4}$$

on the corresponding covariance strata (Drton et al., 2025).

Our numerical study is correspondingly targeted. Rather than benchmark a new inference algorithm, it validates the two theorem-level predictions: local quadratic equivalence of the two excess losses and cancellation of the shared $\lambda \log n$ term in paired free-energy differences across the three one-factor strata.

## 2. Setup and Preliminaries

We fix the loss-based posterior, the local complexity functional, and the Gaussian covariance notation. All asymptotic statements are local.

### 2.1. Generalized posteriors and excess loss

Let $\Theta \subset \mathbb{R}^d$ be the parameter space. Throughout, $U \subset \Theta$ is a fixed compact semianalytic neighborhood of the relevant minimizer fiber, and the prior density $\pi$ is strictly positive and real analytic on a neighborhood of $U$. For a measurable loss $\ell : \mathbb{R}^p \times \Theta \to \mathbb{R}$, define

$$L_n^\ell(\theta) := \sum_{i=1}^n \ell(X_i, \theta), \tag{1}$$

$$\Pi_n^\ell(d\theta) := \frac{\exp\{-L_n^\ell(\theta)\}\pi(\theta)\, d\theta}{\int_\Theta \exp\{-L_n^\ell(u)\}\pi(u)\, du}, \tag{2}$$

$$R_\ell(\theta) := \mathbb{E}_{P_0}[\ell(X, \theta)], \tag{3}$$

$$R_\ell^\star := \inf_{\theta \in \Theta} R_\ell(\theta), \qquad K_\ell(\theta) := R_\ell(\theta) - R_\ell^\star, \tag{4}$$

$$\mathcal{M}_\ell := \operatorname*{argmin}_{\theta \in \Theta} R_\ell(\theta) = \{\theta \in \Theta : K_\ell(\theta) = 0\}. \tag{5}$$

We work at inverse temperature one; a fixed $\beta > 0$ may be absorbed into $\ell$. In singular problems, the primitive local object is the minimizer fiber (5) rather than an identifiable true parameter.

### 2.2. Local zeta functions and learning coefficients

**Definition 1 (Local zeta function and RLCT pair).** *Let $K : U \to [0, \infty)$ be real analytic, with zero set*

$$\mathcal{M} = \{\theta \in U : K(\theta) = 0\}.$$

*Its local zeta function is*

$$\zeta_{K,U}(z) := \int_U K(\theta)^z \pi(\theta)\, d\theta, \qquad \Re z > 0. \tag{6}$$

*By the standard resolution-of-singularities argument, (6) extends meromorphically to $z \in \mathbb{C}$ (Watanabe, 2009). We write*

$$\mathrm{RLCT}_{\mathcal{M}}(K; \pi) = (\lambda_K, m_K)$$

*for the largest pole $-\lambda_K$ and its multiplicity $m_K$.*

**Definition 2 (Local comparison near the zero set).** *For nonnegative functions $K_1, K_2 : U \to [0, \infty)$ with common zero set $\mathcal{M}$, write*

$$K_1 \asymp_{\mathcal{M}} K_2 \tag{7}$$

*if there exist a neighborhood $U_{\mathcal{M}} \subseteq U$ of $\mathcal{M}$ and constants $0 < c \le C < \infty$ such that*

$$c\, K_1(\theta) \le K_2(\theta) \le C\, K_1(\theta), \qquad \theta \in U_{\mathcal{M}}.$$

### 2.3. Gaussian covariance models and factor analysis

Let $\Sigma : \Theta \to \mathbb{S}^p_{++}$ be a real-analytic covariance map and assume $P_0 = N_p(0, \Sigma_0)$ for some $\Sigma_0$ in its image. Up to $\theta$-independent constants, define the Gaussian negative log-loss and the Hyvärinen loss by

$$\ell_{\log}(x, \theta) := \frac{1}{2}\Big( \log\det \Sigma(\theta) + x^\top \Sigma(\theta)^{-1} x \Big), \tag{8}$$

$$\ell_{\mathrm{H}}(x, \theta) := \frac{1}{2} x^\top \Sigma(\theta)^{-2} x - \mathrm{tr}\big(\Sigma(\theta)^{-1}\big). \tag{9}$$

For a target covariance $\Sigma_0 \in \mathbb{S}^p_{++}$, define the covariance fiber

$$\mathcal{M}(\Sigma_0) := \{\theta \in \Theta : \Sigma(\theta) = \Sigma_0\}. \tag{10}$$

In factor analysis,

$$\Sigma(\psi, \Lambda) = \Psi + \Lambda\Lambda^\top, \qquad \Psi = \mathrm{diag}(\psi_1, \dots, \psi_p), \tag{11}$$

with parameter space

$$\Theta^{\mathrm{FA}}_{p,k} = (0, \infty)^p \times \mathbb{R}^{p \times k}, \qquad \mathcal{C}^{\mathrm{FA}}_{p,k} = \{\Psi + \Lambda\Lambda^\top : (\psi, \Lambda) \in \Theta^{\mathrm{FA}}_{p,k}\} \subset \mathbb{S}^p_{++},$$

and fiber

$$\mathcal{M}_{p,k}(\Sigma_0) := \{(\psi, \Lambda) \in \Theta^{\mathrm{FA}}_{p,k} : \Psi + \Lambda\Lambda^\top = \Sigma_0\}. \tag{12}$$

When $k = 1$, we write $a \in \mathbb{R}^p$ and abbreviate $\Sigma(\psi, a) = \mathrm{diag}(\psi) + aa^\top$. Throughout, neighborhoods are chosen so that

$$\underline{\kappa} I_p \preceq \Sigma(\theta) \preceq \overline{\kappa} I_p \qquad \text{for all } \theta \in U,$$

for some $0 < \underline{\kappa} < \overline{\kappa} < \infty$; this makes matrix inversion and spectral comparison locally uniform.

## 3. Universality by Local Comparison

The next results are purely local.

### 3.1. A local comparison principle

**Theorem 3 (Local comparison principle).** *Let $K_1, K_2 : U \to [0,\infty)$ be real analytic with common zero set*

$$\mathcal{M} = \{\theta \in U : K_1(\theta) = 0\} = \{\theta \in U : K_2(\theta) = 0\}.$$

*If $K_1 \asymp_{\mathcal{M}} K_2$, then*

$$\mathrm{RLCT}_{\mathcal{M}}(K_1;\pi) = \mathrm{RLCT}_{\mathcal{M}}(K_2;\pi).$$

*Proof sketch.* Only a neighborhood of $\mathcal{M}$ contributes non-entire behavior to the zeta integral. Take a common analytic resolution $g : Y \to U$ of $K_1$, $K_2$, and $\pi(\theta)\,d\theta$. On each chart with local coordinates $u = (u_1, \ldots, u_r)$,

$$K_j \circ g(u) = a_j(u) \prod_{i=1}^{r} |u_i|^{\alpha_i^{(j)}}, \qquad (\pi \circ g)(u)\,|\det Dg(u)| = b(u) \prod_{i=1}^{r} |u_i|^{\beta_i},$$

with positive analytic units $a_j, b$. Since $K_1 \asymp_{\mathcal{M}} K_2$, the quotient $(K_1 \circ g)/(K_2 \circ g)$ is bounded above and below near $g^{-1}(\mathcal{M})$, forcing $\alpha_i^{(1)} = \alpha_i^{(2)}$ on every chart. The candidate poles and their maximal order therefore coincide for the two local zeta functions, so the largest pole and its multiplicity agree. Full details are given in Appendix B. □

### 3.2. Quadratic-germ universality and sharpness

**Corollary 4 (Quadratic-germ universality).** *Let $h : U \to \mathbb{R}^s$ be real analytic with $\mathcal{M} = h^{-1}(0)$. For $j \in \{1,2\}$, suppose*

$$K_j(\theta) = g_j(h(\theta)),$$

*where $g_j : \mathbb{R}^s \to [0,\infty)$ is real analytic near the origin and satisfies*

$$g_j(0) = 0, \qquad \nabla g_j(0) = 0, \qquad \nabla^2 g_j(0) \succ 0.$$

*Then*

$$\mathrm{RLCT}_{\mathcal{M}}(K_1;\pi) = \mathrm{RLCT}_{\mathcal{M}}(K_2;\pi).$$

*Proof.* Positive definiteness of $\nabla^2 g_j(0)$ gives constants $0 < c_j \le C_j < \infty$ and $\delta > 0$ such that

$$c_j \|u\|_2^2 \le g_j(u) \le C_j \|u\|_2^2, \qquad \|u\|_2 < \delta.$$

After shrinking $U$ so that $h(U) \subset B_\delta(0)$, we obtain $K_j \asymp_{\mathcal{M}} \|h(\theta)\|_2^2$ for $j = 1, 2$. Apply Theorem 3. □

**Proposition 5 (Sharpness under flatter local geometry).** *Let $K_1, K_2 : U \to [0,\infty)$ be real analytic with common zero set $\mathcal{M}$. If, for some integer $q \ge 2$,*

$$K_2 \asymp_{\mathcal{M}} K_1^q,$$

*and $\mathrm{RLCT}_{\mathcal{M}}(K_j;\pi) = (\lambda_j, m_j)$ for $j \in \{1,2\}$, then*

$$\lambda_2 = \lambda_1/q, \qquad m_2 = m_1.$$

*Proof sketch.* On a common analytic resolution, replacing $K_1$ by $K_1^q$ multiplies each vanishing exponent by $q$ and leaves the Jacobian exponents unchanged. Hence every candidate pole is divided by $q$, while the maximal pole order is preserved. Since $K_2 \asymp_{\mathcal{M}} K_1^q$, Theorem 3 transfers the same conclusion to $K_2$. See Appendix C for details. □

Thus quadratic-germ universality is genuinely a quadratic phenomenon: flatter powers rescale the learning coefficient. In particular, the relevant invariant is not whether a loss is likelihood-based, score-based, or decision-theoretic, but whether its population excess risk induces the same identifiable quadratic germ around the minimizer fiber. The next section proves exactly this for Gaussian log-loss and Hyvärinen loss in analytic covariance families.

## 4. Gaussian Covariance Models

We now verify the comparison hypothesis for zero-mean Gaussian covariance families.

### 4.1. Population excess log-loss and Hyvärinen loss

Assume throughout that $P_0 = N_p(0, \Sigma_0)$ and that $\Sigma : \Theta \to \mathbb{S}^p_{++}$ is real analytic with $\Sigma_0$ in its image.

**Proposition 6 (Closed-form excess losses in Gaussian covariance models).** *For every $\theta \in \Theta$,*

$$K_{\log}(\theta) = \frac{1}{2}\Big[\operatorname{tr}\big(\Sigma(\theta)^{-1}\Sigma_0\big) - \log\det\big(\Sigma(\theta)^{-1}\Sigma_0\big) - p\Big], \tag{13}$$

*and*

$$K_{\mathrm{H}}(\theta) = \frac{1}{2}\operatorname{tr}\Big[\Sigma_0\big(\Sigma(\theta)^{-1} - \Sigma_0^{-1}\big)^2\Big] = \frac{1}{2}\Big\|\big(\Sigma(\theta)^{-1} - \Sigma_0^{-1}\big)\Sigma_0^{1/2}\Big\|_{\mathrm{F}}^2. \tag{14}$$

*In particular,*

$$\mathcal{M}_{\ell_{\log}} = \mathcal{M}_{\ell_{\mathrm{H}}} = \mathcal{M}(\Sigma_0) = \{\theta \in \Theta : \Sigma(\theta) = \Sigma_0\}.$$

*Proof.* Using $\mathbb{E}_{P_0}[XX^\top] = \Sigma_0$,

$$R_{\ell_{\log}}(\theta) = \frac{1}{2}\Big(\log\det\Sigma(\theta) + \operatorname{tr}\big(\Sigma(\theta)^{-1}\Sigma_0\big)\Big),$$

so subtracting its value at $\Sigma_0$ gives (13). Likewise,

$$R_{\ell_{\mathrm{H}}}(\theta) = \frac{1}{2}\operatorname{tr}\big(\Sigma(\theta)^{-2}\Sigma_0\big) - \operatorname{tr}\big(\Sigma(\theta)^{-1}\big),$$

and subtracting the value at any $\theta^\star \in \mathcal{M}(\Sigma_0)$ yields (14). The zero-set claim follows from strict positivity of $t \mapsto t - \log t - 1$ on $(0,\infty)\setminus\{1\}$ and from (14). □

### 4.2. Local quadratic equivalence

**Theorem 7 (Local quadratic equivalence of Gaussian log-loss and Hyvärinen loss).** *Let $U \subset \Theta$ be compact. Then there exist constants $0 < c_U \le C_U < \infty$ such that, for all $\theta \in U$,*

$$c_U\ \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2 \le K_{\log}(\theta) \le C_U\ \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2, \tag{15}$$

*and*

$$c_U\ \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2 \le K_{\mathrm{H}}(\theta) \le C_U\ \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2. \tag{16}$$

*Hence*

$$K_{\log} \asymp_{\mathcal{M}(\Sigma_0)} K_{\mathrm{H}}.$$

*Proof sketch.* Compactness of $U$ and continuity of $\Sigma$ give

$$\underline{\kappa} I_p \preceq \Sigma(\theta) \preceq \overline{\kappa} I_p, \qquad \theta \in U,$$

for some $0 < \underline{\kappa} \le \overline{\kappa} < \infty$. Hence matrix inversion is bi-Lipschitz on $\Sigma(U)$:

$$\big\|\Sigma(\theta)^{-1} - \Sigma_0^{-1}\big\|_{\mathrm{F}} \asymp \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}, \qquad \theta \in U, \tag{17}$$

using

$$\Sigma(\theta)^{-1} - \Sigma_0^{-1} = \Sigma(\theta)^{-1}(\Sigma_0 - \Sigma(\theta))\Sigma_0^{-1}, \qquad \Sigma(\theta) - \Sigma_0 = \Sigma(\theta)(\Sigma_0^{-1} - \Sigma(\theta)^{-1})\Sigma_0.$$

For the Hyvärinen loss, (14) gives

$$K_{\mathrm{H}}(\theta) \asymp \left\|\Sigma(\theta)^{-1} - \Sigma_0^{-1}\right\|_{\mathrm{F}}^2 \asymp \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2 .$$

For the log-loss, let

$$B(\theta) := \Sigma_0^{1/2}\Sigma(\theta)^{-1}\Sigma_0^{1/2},$$

with eigenvalues $\mu_1(\theta), \ldots, \mu_p(\theta)$. Then (13) becomes

$$K_{\log}(\theta) = \frac{1}{2}\sum_{i=1}^{p} \phi(\mu_i(\theta)), \qquad \phi(t) := t - \log t - 1.$$

Because the spectrum of $B(\theta)$ stays in a compact subset of $(0, \infty)$, one has $\phi(t) \asymp (t-1)^2$ there, so

$$K_{\log}(\theta) \asymp \|B(\theta) - I_p\|_{\mathrm{F}}^2 .$$

Since

$$B(\theta) - I_p = \Sigma_0^{1/2}\big(\Sigma(\theta)^{-1} - \Sigma_0^{-1}\big)\Sigma_0^{1/2},$$

conjugation by the fixed matrix $\Sigma_0^{1/2}$ yields

$$\|B(\theta) - I_p\|_{\mathrm{F}} \asymp \left\|\Sigma(\theta)^{-1} - \Sigma_0^{-1}\right\|_{\mathrm{F}} .$$

Combining this with (17) proves (15)–(16). Details are collected in Appendix D. □

### 4.3. Universality in analytic covariance families

**Corollary 8 (Universality in analytic Gaussian covariance families).** *Let $\Sigma : \Theta \to \mathbb{S}_{++}^p$ be real analytic, and let $\Sigma_0$ lie in its image. Then*

$$\mathrm{RLCT}_{\mathcal{M}(\Sigma_0)}(K_{\log}; \pi) = \mathrm{RLCT}_{\mathcal{M}(\Sigma_0)}(K_{\mathrm{H}}; \pi).$$

*Proof.* Set $h(\theta) = \mathrm{vech}\big(\Sigma(\theta) - \Sigma_0\big)$. Then $h$ is real analytic, $h^{-1}(0) = \mathcal{M}(\Sigma_0)$, and

$$\|h(\theta)\|_2^2 \asymp \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2 .$$

By Theorem 7, both excess losses are locally equivalent to $\|h(\theta)\|_2^2$, so Theorem 3 gives the claim. □

## 5. Exact Transfer to Gaussian Factor Analysis

Since the factor-analysis covariance map is real analytic, Corollary 8 applies directly.

### 5.1. Exact transfer theorem

**Theorem 9 (Exact transfer to Gaussian factor analysis).** *For every $p \geq 1$, $k \geq 0$, and every $\Sigma_0 \in \mathcal{C}_{p,k}^{\mathrm{FA}}$,*

$$\mathrm{RLCT}_{\mathcal{M}_{p,k}(\Sigma_0)}(K_{\log}; \pi) = \mathrm{RLCT}_{\mathcal{M}_{p,k}(\Sigma_0)}(K_{\mathrm{H}}; \pi).$$

*Proof.* Apply Corollary 8 with $\Theta = \Theta_{p,k}^{\mathrm{FA}}$ and $\mathcal{M}(\Sigma_0) = \mathcal{M}_{p,k}(\Sigma_0)$. □

Write $d_s := (s+1)p - \binom{s}{2}$, and let $r(\Sigma_0)$ denote the minimal latent rank realizing $\Sigma_0$.

**Corollary 10 (Transferred factor-analysis results).** *For every* $\Sigma_0 \in \mathcal{C}^{\mathrm{FA}}_{p,k}$,

$$\lambda^{\mathrm{H}}_{p,k}(\Sigma_0) = \lambda^{\log}_{p,k}(\Sigma_0).$$

*Hence*

$$\lambda^{\mathrm{H}}_{p,k}(\Sigma_0) \le \frac{p(k+2) + r(\Sigma_0)\big(p-k+1\big)}{4}, \qquad d_{r(\Sigma_0)} \le p(p+1)/2. \tag{18}$$

*Moreover, the exact Bayes-side formulas of Drton et al. (2025) transfer unchanged. In particular:*

1. *If* $\Sigma_0$ *is diagonal, then*

$$\lambda^{\mathrm{H}}_{p,k}(\Sigma_0) = \begin{cases} \dfrac{p(k+2)}{4}, & k \le p-1, \\ \dfrac{p^2+p+1}{4}, & k = p, \end{cases}$$

*and*

$$m^{\mathrm{H}}_{p,k}(\Sigma_0) = \begin{cases} 1, & k \le p-2, \ \text{or } k = p, \ \text{or } (p,k) = (1,0), \\ p-1, & k = p-1 > 0. \end{cases}$$

2. *If* $\Sigma_0$ *is generic of latent rank one, then*

$$\lambda^{\mathrm{H}}_{p,k}(\Sigma_0) = \begin{cases} \dfrac{pk+3p-k+1}{4}, & k \le p-2, \\ \dfrac{p(p+1)}{4}, & k \in \{p-1, p\}, \end{cases} \qquad m^{\mathrm{H}}_{p,k}(\Sigma_0) = 1.$$

3. *If* $\Sigma_0$ *is generic of latent rank* $k$ *and* $d_k \le p(p+1)/2$, *then*

$$\lambda^{\mathrm{H}}_{p,k}(\Sigma_0) = \frac{d_k}{2}, \qquad m^{\mathrm{H}}_{p,k}(\Sigma_0) = 1.$$

*Proof.* The identity of learning coefficients follows from Theorem 9. The bound (18) and the stated exact formulas are then inherited directly from the corresponding ordinary-Gaussian results in Drton et al. (2025). □

### 5.2. Exact stratification of the one-factor model

**Corollary 11 (Exact one-factor transfer).** *Suppose* $k = 1$ *and let* $\Sigma_0 = \mathrm{diag}(\psi_0) + a_0 a_0^\top$. *Then*

$$\big(\lambda^{\mathrm{H}}_{p,1}(\Sigma_0), m^{\mathrm{H}}_{p,1}(\Sigma_0)\big) = \begin{cases} \big(p, 1\big), & \#\{i : a_{0i} \neq 0\} > 2, \\ \big((2p-1)/2, 1\big), & \#\{i : a_{0i} \neq 0\} = 2, \\ \big(3p/4, 1\big), & \#\{i : a_{0i} \neq 0\} \le 1. \end{cases}$$

*The middle case is equivalent to* $\Sigma_0$ *having exactly one nonzero off-diagonal pair, and the last to* $\Sigma_0$ *being diagonal.*

*Proof.* The corresponding ordinary Gaussian values are exactly the one-factor results of Drton et al. (2025); Theorem 9 then yields the Hyvärinen values. □

These three strata provide the natural numerical test bed, since they give distinct explicit coefficients within the same latent-variable model.

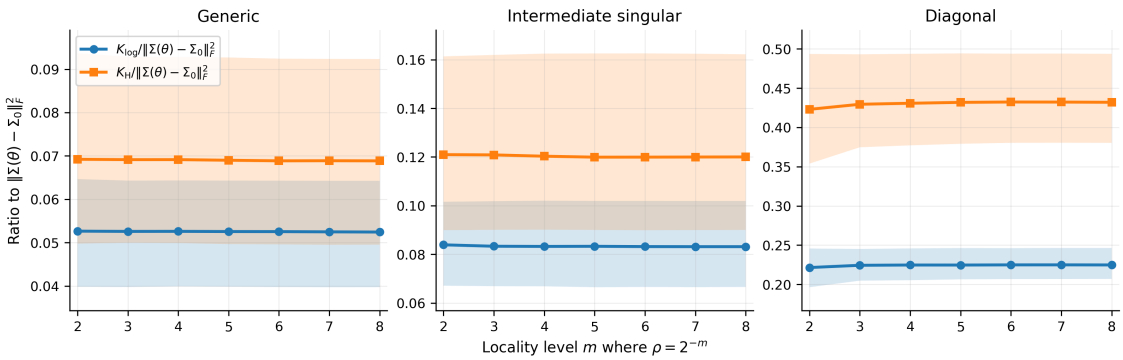


**Figure 1. Local quadratic equivalence.** For perturbations $(\eta_\rho, a_\rho) = (\eta_0, a_0) + \rho u$ with $\rho = 2^{-m}$, we plot $K_{\log}(\theta_\rho)/\|\Sigma(\theta_\rho) - \Sigma_0\|_{\mathrm{F}}^2$ and $K_{\mathrm{H}}(\theta_\rho)/\|\Sigma(\theta_\rho) - \Sigma_0\|_{\mathrm{F}}^2$. Solid lines are medians over random directions and bands are interquartile ranges. Across all three strata, both ratios stay bounded away from zero and infinity in the local regime, in agreement with Theorem 7.

## 6. Numerical Validation

Our numerical study targets the two theorem-level consequences proved above: local quadratic equivalence of the excess losses and equality of the $\lambda \log n$ coefficient.

**Design.** We work in the one-factor model with $p = 4$, $\psi_0 = (1.0, 1.2, 0.9, 1.1)$, and

$$a_0 \in \{(1.1, 0.9, 1.0, 0.8),\ (1, 1, 0, 0),\ (0, 0, 0, 0)\},$$

corresponding to the generic, intermediate, and diagonal strata in Corollary 11. The target covariance is $\Sigma_0 = \mathrm{diag}(\psi_0) + a_0 a_0^\top$, data are sampled i.i.d. from $N_4(0, \Sigma_0)$, and the three exact coefficients are $\lambda_\star = 4$, 3.5, and 3, respectively.

**Local quadratic equivalence.** For Theorem 7, we evaluate the closed-form excess losses in Proposition 6 on perturbations of the target parameter. Writing $\eta = \log \psi$, we set $(\eta_\rho, a_\rho) = (\eta_0, a_0) + \rho u$ with $\rho = 2^{-m}$, $m = 2, \ldots, 8$, and random unit directions $u$. For each perturbation we compute

$$\frac{K_{\log}(\theta_\rho)}{\|\Sigma(\theta_\rho) - \Sigma_0\|_{\mathrm{F}}^2} \quad \text{and} \quad \frac{K_{\mathrm{H}}(\theta_\rho)}{\|\Sigma(\theta_\rho) - \Sigma_0\|_{\mathrm{F}}^2}.$$

Figure 1 shows that both ratios stabilize to positive finite constants in all three strata, which is exactly the numerical signature of Theorem 7.

**Transferred $\log n$ coefficient.** To test Theorem 9 directly, we estimate the centered free energy under both losses using adaptive tempered SMC (Del Moral et al., 2006); implementation details are given in Appendix G. The two losses are evaluated on the same synthetic datasets, and repeated SMC estimates are aggregated on the normalizing-constant scale before taking logarithms. This pairing is important: it removes most dataset-level variation and isolates the specific asymptotic term that the theory predicts should cancel. Denote the paired difference by

$$D_n := \widetilde{F}_n^{\mathrm{H}} - \widetilde{F}_n^{\log}.$$

If the local RLCT pairs agree, then the $\lambda \log n$ term cancels and $D_n$ should be $O_p(1)$. We therefore fit

$$D_n = \alpha + \beta \log n + \varepsilon_n$$

over the asymptotic regime $n \geq 1600$ by weighted least squares with dataset-level bootstrap uncertainties. The fitted slopes are statistically indistinguishable from zero in all three strata; see Fig. 2. This is the direct numerical signature of exact RLCT transfer.

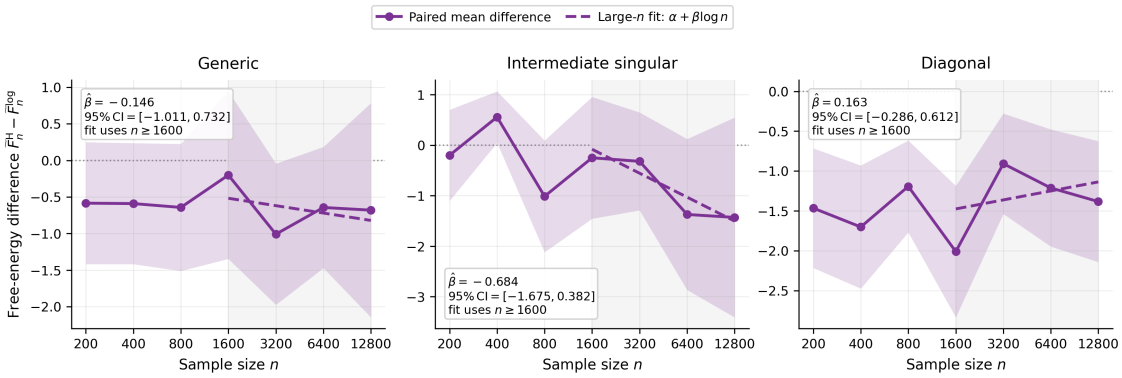


**Figure 2. Paired free-energy difference.** The plotted quantity is $\widetilde{F}_n^{\mathrm{H}} - \widetilde{F}_n^{\log}$ on shared synthetic datasets. Dashed lines are weighted least-squares fits $\alpha + \beta \log n$ over $n \geq 1600$; bands are pointwise 95% intervals across datasets. In all three strata the fitted slopes are near zero, consistent with cancellation of the common $\lambda \log n$ term.

Additional diagnostics are reported in Appendix H; they support the same conclusion but are not needed for the main claim.

## 7. Discussion

We established a local transfer principle for generalized Bayes losses. The abstract result is that locally comparable nonnegative analytic excess-loss germs have the same RLCT pair. In Gaussian covariance models, both the Gaussian excess log-loss and the excess Hyvärinen loss are locally equivalent to the squared covariance error $\|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2$. This yields exact transfer of local singular complexity from ordinary Gaussian Bayes to Hyvärinen generalized Bayes throughout factor analysis, including the explicit one-factor coefficients.

The result is intentionally local. It identifies the leading singular-complexity term, not global posterior equivalence, Bayes-factor semantics, or the $O_p(1)$ remainder; the latter can differ across losses, as seen in the free-energy offsets. This distinction matters conceptually. Outside negative log-likelihood, exponential tilting is best interpreted in the general Bayesian framework of Bissiri et al. (2016), and recent work emphasizes that ordinary Bayesian belief updating is exceptional rather than generic in that class (McAlinn and Takanashi, 2026). Our theorem therefore concerns the geometry of the excess-risk germ and the resulting $\lambda \log n$ term, not a full identification of the two posteriors.

The proof strategy also suggests a broader research program. First, one would like to know how far the same local-comparison mechanism extends beyond Gaussian covariance families to other singular latent-variable models. Second, Hyvärinen loss should be only one representative of a wider class of smooth scores whose excess risks induce the same quadratic identifiable germ. Third, once RLCT transfer is understood, the next issue is lower-order structure: multiplicities in broader non-generic strata, refined $O_p(1)$ expansions, and whether score-based losses preserve more detailed asymptotic invariants than the learning coefficient alone. The present paper isolates the basic mechanism needed for such extensions: identify the minimizer fiber, prove local comparison to a common quadratic germ, and transfer singular complexity without redoing the full algebraic analysis for each loss.

## References

Pier Giovanni Bissiri, Chris C. Holmes, and Stephen G. Walker. A general framework for updating belief distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*,

78(5):1103–1130, 2016. doi: 10.1111/rssb.12158.

Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.

Mathias Drton, Elizabeth Gross, Dimitra Kosta, Anton Leykin, Seth Sullivant, and Daniel Windisch. Singular learning theory for factor analysis, 2025.

Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6:695–709, 2005.

Jack Jewson and David Rossell. General bayesian loss function selection and the use of improper models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84(5):1640–1665, 2022. doi: 10.1111/rssb.12553.

Kenichiro McAlinn and Kōsaku Takanashi. When is generalized bayes bayesian? a decision-theoretic characterization of loss-based updating, 2026.

Sumio Watanabe. *Algebraic Geometry and Statistical Learning Theory*. Cambridge University Press, 2009.

Sumio Watanabe. A widely applicable bayesian information criterion. *Journal of Machine Learning Research*, 14(27):867–897, 2013.

## A. Analytic Setup and Local Zeta Preliminaries

This appendix records the analytic background used throughout the proof appendices. Nothing here is new: the underlying framework is standard singular learning theory in the sense of Watanabe (2009, 2013). Our goal is simply to isolate the local zeta-function conventions and the two elementary facts that will be used repeatedly later: first, meromorphic continuation of the local zeta function in the analytic setting; second, the fact that the RLCT pair depends only on the germ of the excess-loss function near its zero set.

### A.1. Standing analytic assumptions

We work on a fixed compact semianalytic neighborhood $U \subset \Theta \subset \mathbb{R}^d$, where $\Theta$ is open and $U$ contains the relevant minimizer set in its interior. Let $\Omega \subset \Theta$ be an open neighborhood of $U$. The standing assumptions are:

1. $K : \Omega \to [0, \infty)$ is real analytic and not identically zero.
2. The local zero set
$$\mathcal{M} := \{\theta \in U : K(\theta) = 0\}$$
is nonempty and compact.
3. The prior density $\pi : \Omega \to (0, \infty)$ is real analytic and strictly positive on $\Omega$.

These hypotheses are exactly those needed for the standard singular-learning-theory construction of the local zeta function. In particular, the pair $(K, \pi)$ is analytic near $\mathcal{M}$, and all singular behavior arises from the vanishing of $K$ along $\mathcal{M}$, not from the prior density.

**Definition 12 (Local zeta function and local RLCT).** *For $\Re z > 0$, define*

$$\zeta_{K,U}(z) := \int_U K(\theta)^z \pi(\theta)\, d\theta. \tag{19}$$

*In the analytic setting above, $\zeta_{K,U}$ admits a meromorphic continuation to $z \in \mathbb{C}$. The local RLCT pair of $K$ along $\mathcal{M}$, relative to $\pi$, is denoted*

$$\mathrm{RLCT}_{\mathcal{M}}(K;\pi) = (\lambda_K, m_K),$$

*where $-\lambda_K$ is the pole of $\zeta_{K,U}$ with largest real part and $m_K$ is its multiplicity.*

The terminology "local" is essential: by construction, $\lambda_K$ and $m_K$ are intended to describe only the singular contribution coming from the germ of $K$ near $\mathcal{M}$.

### A.2. Standard meromorphic continuation

The existence and basic pole structure of $\zeta_{K,U}$ follow from resolution of singularities. We use this result as a black box.

**Theorem 13 (Standard meromorphic continuation).** *Under the standing assumptions above, the function $\zeta_{K,U}$ defined initially for $\Re z > 0$ extends meromorphically to $z \in \mathbb{C}$. Moreover, its poles are negative rational numbers of finite multiplicity, and there is a unique pole with maximal real part. Equivalently, the local RLCT pair $(\lambda_K, m_K)$ is well defined.*

*Proof.* This is standard in singular learning theory. A real-analytic resolution of singularities reduces $K$ to normal-crossing form

$$K \circ g(u) = a(u) \prod_{i=1}^{r} u_i^{2k_i},$$

with $a(u)$ a positive analytic unit, and simultaneously monomializes the pullback of the volume form $\pi(\theta)\, d\theta$; see Watanabe (2009, Chs. 6–7). Substituting the resolved form into (19) yields a finite sum of products of one-dimensional Mellin-type integrals, each of which extends meromorphically with rational poles. The stated consequences follow immediately. □

We will not use the explicit resolved form in the sequel; the only properties needed later are meromorphic continuation and locality.

### A.3. Locality with respect to the zero set

The next lemma shows that removing any region that stays a positive distance away from the zero set changes the zeta function only by an entire term. This is the basic reason the RLCT depends only on the local germ near $\mathcal{M}$.

**Lemma 14 (The complement of a neighborhood of $\mathcal{M}$ contributes an entire term).** *Let $W \subset U$ be an open neighborhood of $\mathcal{M}$ with compact closure $\overline{W} \subset U$. Then the function*

$$H_{K,U\setminus W}(z) := \int_{U\setminus W} K(\theta)^z \pi(\theta)\, d\theta \tag{20}$$

*extends to an entire function on $\mathbb{C}$.*

*Proof.* Since $K$ is continuous, nonnegative, and vanishes exactly on $\mathcal{M}$, while the compact set $U \setminus W$ is disjoint from $\mathcal{M}$, there exists a constant $c_W > 0$ such that

$$K(\theta) \geq c_W > 0, \qquad \theta \in U \setminus W.$$

Likewise, compactness of $U \setminus W$ gives a finite upper bound

$$K(\theta) \leq C_W < \infty, \qquad \theta \in U \setminus W.$$

Hence the real logarithm $\log K(\theta)$ is well defined and bounded on $U \setminus W$.

For $z \in \mathbb{C}$, define

$$K(\theta)^z := \exp\big(z \log K(\theta)\big), \qquad \theta \in U \setminus W.$$

Then for each fixed $\theta \in U \setminus W$, the map

$$z \longmapsto K(\theta)^z \pi(\theta)$$

is entire. Moreover, if $B \subset \mathbb{C}$ is compact, then

$$\sup_{z \in B,\ \theta \in U \setminus W} \big|K(\theta)^z \pi(\theta)\big| \leq \exp\Big( \sup_{z \in B} |z| \cdot \sup_{\theta \in U \setminus W} |\log K(\theta)| \Big) \sup_{\theta \in U \setminus W} \pi(\theta) < \infty.$$

Thus the integrand is uniformly bounded on compact subsets of $\mathbb{C}$, and one may differentiate under the integral sign any number of times. In particular,

$$\frac{d^m}{dz^m} H_{K,U \setminus W}(z) = \int_{U \setminus W} (\log K(\theta))^m K(\theta)^z \pi(\theta)\, d\theta, \qquad m \geq 0,$$

so $H_{K,U \setminus W}$ is entire. □

The promised locality statement is now immediate.

**Corollary 15 (Independence of the neighborhood).** *Let $W \subset U$ be an open neighborhood of $\mathcal{M}$ with compact closure in $U$. Then*

$$\mathrm{RLCT}_{\mathcal{M}}(K; \pi)$$

*computed from the zeta integral over $U$ is identical to the RLCT pair computed from the zeta integral over $W$. Equivalently,*

$$\zeta_{K,U}(z) - \zeta_{K,W}(z) \tag{21}$$

*is an entire function, so both zeta functions have the same pole of maximal real part and the same multiplicity there.*

*Proof.* Decompose

$$\zeta_{K,U}(z) = \zeta_{K,W}(z) + H_{K,U \setminus W}(z),$$

where $H_{K,U \setminus W}$ is defined in (20). By Lemma 14, the second term is entire. Adding or subtracting an entire function does not alter the pole set of a meromorphic function, and in particular does not alter the pole of maximal real part or its multiplicity. Hence the local RLCT pair is the same whether computed on $U$ or on $W$. □

**Remark 16 (Dependence only on the germ near $\mathcal{M}$).** *The preceding corollary is the precise sense in which the RLCT is local. Any modification of $K$ or $\pi$ outside an arbitrarily small neighborhood of $\mathcal{M}$ changes the local zeta function only by an entire term and therefore leaves $(\lambda_K, m_K)$ unchanged. This locality is used repeatedly in later appendices, especially in the comparison argument of Appendix B.*

## B. Proof of the Local Comparison Theorem

This appendix proves Theorem 3. We use the standard Laplace characterization of the local RLCT from singular learning theory (Watanabe, 2009, 2013). This route has one technical advantage: once the leading Laplace asymptotic is known, local comparability of two excess-loss germs immediately yields a two-sided exponential sandwich, and hence equality of the corresponding RLCT pairs. In particular, the proof avoids having to track possible cancellation among resolution charts.

Throughout, $U \subset \Theta$ is the fixed compact semianalytic neighborhood from Sections 2.1 and 2.2. Let $K : U \to [0, \infty)$ be real analytic, let

$$\mathcal{M} = \{\theta \in U : K(\theta) = 0\},$$

and write

$$\mathcal{L}_K(t; A) := \int_A e^{-tK(\theta)} \pi(\theta)\, d\theta, \qquad t > 0, \tag{22}$$

for the Laplace integral over a measurable set $A \subseteq U$.

### B.1. Localization of the Laplace integral

The first observation is that, for purposes of the RLCT, only a neighborhood of the zero set matters.

**Lemma 17 (Exponential negligibility away from the zero set).** *Let $W \subset U$ be an open neighborhood of $\mathcal{M}$ with compact closure $\overline{W} \subset U$. Then there exists $c_W > 0$ such that*

$$\mathcal{L}_K(t; U \setminus W) \le e^{-c_W t} \int_{U \setminus W} \pi(\theta)\, d\theta, \qquad t > 0. \tag{23}$$

*In particular,*

$$\mathcal{L}_K(t; U \setminus W) = O(e^{-c_W t}) \qquad \text{as } t \to \infty.$$

*Proof.* Since $K$ is continuous, nonnegative, and vanishes exactly on $\mathcal{M}$, while $U \setminus W$ is compact and disjoint from $\mathcal{M}$, the minimum

$$c_W := \inf_{\theta \in U \setminus W} K(\theta)$$

is strictly positive. Therefore, for every $\theta \in U \setminus W$,

$$e^{-tK(\theta)} \le e^{-c_W t}.$$

Multiplying by $\pi(\theta)$ and integrating over $U \setminus W$ proves (23). □

### B.2. Laplace asymptotics and the RLCT

The next theorem is standard in singular learning theory. It identifies the RLCT pair with the leading polynomial-logarithmic asymptotic of the Laplace integral.

**Theorem 18 (Standard RLCT–Laplace correspondence).** *Let $K : U \to [0, \infty)$ be real analytic, and let*

$$\mathrm{RLCT}_{\mathcal{M}}(K; \pi) = (\lambda_K, m_K)$$

*be the local RLCT pair from Definition 1. Then there exists a constant $C_K \in (0, \infty)$ such that*

$$\mathcal{L}_K(t; U) = C_K\, t^{-\lambda_K} (\log t)^{m_K - 1} \big(1 + o(1)\big) \qquad \text{as } t \to \infty. \tag{24}$$

*Moreover, for every open neighborhood $W \subset U$ of $\mathcal{M}$ with compact closure in $U$,*

$$\mathcal{L}_K(t; W) = C_K\, t^{-\lambda_K} (\log t)^{m_K - 1} \big(1 + o(1)\big) \qquad \text{as } t \to \infty. \tag{25}$$

*Proof.* The global statement (24) is standard; see Watanabe (2009, Ch. 6) and Watanabe (2013). To obtain the local version (25), write

$$\mathcal{L}_K(t;U) = \mathcal{L}_K(t;W) + \mathcal{L}_K(t;U \setminus W).$$

By Lemma 17, the second term is exponentially small in $t$, whereas the first term has polynomial-logarithmic order. Hence the exponentially small remainder is negligible relative to $t^{-\lambda_K}(\log t)^{m_K-1}$, and (25) follows. □

The following elementary comparison lemma will be used repeatedly.

**Lemma 19 (Comparison of polynomial-logarithmic asymptotics).** *Let $f, g : (0,\infty) \to (0,\infty)$ be positive functions satisfying*

$$f(t) = a\,t^{-\lambda}(\log t)^{m-1}\big(1+o(1)\big), \qquad g(t) = b\,t^{-\mu}(\log t)^{n-1}\big(1+o(1)\big),$$

*for some $a, b > 0$, $\lambda, \mu > 0$, and integers $m, n \geq 1$. If there exist constants $0 < c \leq C < \infty$ such that*

$$c\,f(t) \leq g(t) \leq C\,f(t)$$

*for all sufficiently large $t$, then*

$$\lambda = \mu \qquad \text{and} \qquad m = n.$$

*Proof.* Consider the ratio

$$\frac{g(t)}{f(t)} = \frac{b}{a}\,t^{-(\mu-\lambda)}(\log t)^{n-m}\big(1+o(1)\big).$$

By assumption, this ratio remains bounded above and below by positive constants for all sufficiently large $t$. If $\mu \neq \lambda$, then the factor $t^{-(\mu-\lambda)}$ tends either to 0 or to $\infty$, which is impossible. Hence $\mu = \lambda$. With this identity fixed, the ratio reduces to

$$\frac{g(t)}{f(t)} = \frac{b}{a}(\log t)^{n-m}\big(1+o(1)\big).$$

If $n \neq m$, then $(\log t)^{n-m}$ again tends either to 0 or to $\infty$, contradicting boundedness away from both 0 and $\infty$. Therefore $n = m$. □

We will also use the trivial fact that replacing $t$ by a positive constant multiple does not change the leading polynomial-logarithmic scale.

**Lemma 20 (Rescaling the Laplace parameter).** *Let*

$$f(t) = a\,t^{-\lambda}(\log t)^{m-1}\big(1+o(1)\big) \qquad \text{as } t \to \infty,$$

*with $a > 0$, $\lambda > 0$, and $m \geq 1$. Then for every $c > 0$,*

$$f(ct) = a\,c^{-\lambda}t^{-\lambda}(\log t)^{m-1}\big(1+o(1)\big) \qquad \text{as } t \to \infty.$$

*Proof.* Since

$$\log(ct) = \log t + \log c = \log t\,\big(1+o(1)\big) \qquad \text{as } t \to \infty,$$

we have

$$(\log(ct))^{m-1} = (\log t)^{m-1}\big(1+o(1)\big).$$

Substituting this into the asymptotic formula for $f(ct)$ proves the claim. □

### B.3. Proof of Theorem 3

We can now prove the local comparison theorem.

*Proof of Theorem 3.* Let $K_1, K_2 : U \to [0, \infty)$ be real analytic with common zero set

$$\mathcal{M} = \{\theta \in U : K_1(\theta) = 0\} = \{\theta \in U : K_2(\theta) = 0\},$$

and suppose

$$K_1 \asymp_{\mathcal{M}} K_2.$$

By Definition 2, there exist an open neighborhood $W \subset U$ of $\mathcal{M}$ with compact closure in $U$ and constants $0 < c \leq C < \infty$ such that

$$c\, K_1(\theta) \leq K_2(\theta) \leq C\, K_1(\theta), \qquad \theta \in W. \tag{26}$$

For $j \in \{1, 2\}$, define

$$\mathcal{L}_j(t) := \mathcal{L}_{K_j}(t; W) = \int_W e^{-tK_j(\theta)} \pi(\theta)\, d\theta.$$

By Theorem 18, we have

$$\mathcal{L}_j(t) = C_j\, t^{-\lambda_j} (\log t)^{m_j - 1} \big(1 + o(1)\big), \qquad t \to \infty, \tag{27}$$

where

$$(\lambda_j, m_j) = \mathrm{RLCT}_{\mathcal{M}}(K_j; \pi), \qquad C_j > 0.$$

Now exponentiate the bounds in (26). For every $\theta \in W$ and every $t > 0$,

$$e^{-CtK_1(\theta)} \leq e^{-tK_2(\theta)} \leq e^{-ctK_1(\theta)}.$$

Integrating over $W$ against $\pi(\theta)\, d\theta$ yields

$$\mathcal{L}_1(Ct) \leq \mathcal{L}_2(t) \leq \mathcal{L}_1(ct), \qquad t > 0. \tag{28}$$

Applying Lemma 20 to the asymptotic expansion of $\mathcal{L}_1$, we obtain

$$\mathcal{L}_1(Ct) = C_1 C^{-\lambda_1} t^{-\lambda_1} (\log t)^{m_1 - 1} \big(1 + o(1)\big)$$

and

$$\mathcal{L}_1(ct) = C_1 c^{-\lambda_1} t^{-\lambda_1} (\log t)^{m_1 - 1} \big(1 + o(1)\big).$$

Hence (28) implies that $\mathcal{L}_2(t)$ is bounded above and below by positive multiples of

$$t^{-\lambda_1} (\log t)^{m_1 - 1}$$

for all sufficiently large $t$. Since $\mathcal{L}_2(t)$ itself has the asymptotic form (27), Lemma 19 yields

$$\lambda_2 = \lambda_1 \qquad \text{and} \qquad m_2 = m_1.$$

Therefore

$$\mathrm{RLCT}_{\mathcal{M}}(K_1; \pi) = \mathrm{RLCT}_{\mathcal{M}}(K_2; \pi),$$

which is exactly the claim of Theorem 3. □

The theorem has an immediate and useful special case.

**Corollary 21 (Invariance under positive analytic units).** *Let $K : U \to [0, \infty)$ be real analytic with zero set $\mathcal{M}$, and let $u : U \to (0, \infty)$ be a positive real-analytic function. Then*

$$\mathrm{RLCT}_{\mathcal{M}}(K; \pi) = \mathrm{RLCT}_{\mathcal{M}}(uK; \pi).$$

*Proof.* Since $u$ is positive and continuous on the compact neighborhood $U$, there exist constants $0 < c \le C < \infty$ such that

$$c \le u(\theta) \le C, \qquad \theta \in U.$$

Hence

$$c\,K(\theta) \le u(\theta)K(\theta) \le C\,K(\theta), \qquad \theta \in U,$$

so $K \asymp_{\mathcal{M}} uK$. The claim follows from Theorem 3. □

## C. Quadratic-Germ Universality and Sharpness

This appendix proves Corollary 4 and Proposition 5. The first result identifies the stable universality class relevant for the paper: if two analytic excess-loss germs factor through the same local minimizer map and both have nondegenerate quadratic behavior in the induced identifiable coordinates, then they have the same local RLCT pair. The second result shows that this quadratic condition is sharp: replacing a quadratic germ by a flatter power changes the learning coefficient by the reciprocal power. The underlying zeta-function viewpoint is standard in singular learning theory (Watanabe, 2009, 2013).

### C.1. Quadratic-germ universality

We begin with the elementary analytic fact that a nondegenerate quadratic germ is locally equivalent to the Euclidean square norm.

**Lemma 22 (Quadratic sandwich for analytic germs).** *Let $g : V \to \mathbb{R}$ be real analytic on a neighborhood $V \subset \mathbb{R}^s$ of the origin. Assume*

$$g(0) = 0, \qquad \nabla g(0) = 0, \qquad \nabla^2 g(0) \succ 0.$$

*Then there exist constants $0 < c_g \le C_g < \infty$ and $\delta_g > 0$ such that*

$$c_g \|u\|_2^2 \le g(u) \le C_g \|u\|_2^2, \qquad \|u\|_2 < \delta_g. \tag{29}$$

*Proof.* Since $\nabla^2 g(0)$ is positive definite and $\nabla^2 g$ is continuous, there exist constants $0 < m_g \le M_g < \infty$ and $\delta_g > 0$ such that

$$m_g I_s \preceq \nabla^2 g(v) \preceq M_g I_s, \qquad \|v\|_2 < \delta_g.$$

Fix $u \in \mathbb{R}^s$ with $\|u\|_2 < \delta_g$. By Taylor's theorem with integral remainder and the assumptions $g(0) = 0$, $\nabla g(0) = 0$,

$$g(u) = \int_0^1 (1-t)\, u^\top \nabla^2 g(tu)\, u\, dt.$$

Therefore

$$\int_0^1 (1-t)\, m_g \|u\|_2^2 \, dt \le g(u) \le \int_0^1 (1-t)\, M_g \|u\|_2^2 \, dt.$$

Since $\int_0^1 (1-t)\, dt = 1/2$, we obtain

$$\frac{m_g}{2} \|u\|_2^2 \le g(u) \le \frac{M_g}{2} \|u\|_2^2.$$

Thus (29) holds with $c_g = m_g/2$ and $C_g = M_g/2$. □

We now prove the structured universality statement from Section 3.

*Proof of Corollary 4.* Let $h : U \to \mathbb{R}^s$ be real analytic with $\mathcal{M} = h^{-1}(0)$, and suppose

$$K_j(\theta) = g_j(h(\theta)), \qquad j \in \{1, 2\},$$

where each $g_j$ is real analytic near the origin and satisfies

$$g_j(0) = 0, \qquad \nabla g_j(0) = 0, \qquad \nabla^2 g_j(0) \succ 0.$$

Applying Lemma 22 to each $g_j$, we obtain constants $0 < c_j \leq C_j < \infty$ and $\delta_j > 0$ such that

$$c_j \|u\|_2^2 \leq g_j(u) \leq C_j \|u\|_2^2, \qquad \|u\|_2 < \delta_j.$$

Set $\delta := \min(\delta_1, \delta_2)$.

Because $h$ is continuous and $h(\theta) = 0$ for every $\theta \in \mathcal{M}$, there exists an open neighborhood $W \subset U$ of $\mathcal{M}$ such that

$$\|h(\theta)\|_2 < \delta, \qquad \theta \in W.$$

Hence, for every $\theta \in W$,

$$c_j \|h(\theta)\|_2^2 \leq K_j(\theta) \leq C_j \|h(\theta)\|_2^2, \qquad j \in \{1, 2\}.$$

In particular,

$$K_1 \asymp_{\mathcal{M}} \|h(\theta)\|_2^2 \asymp_{\mathcal{M}} K_2.$$

Since $\{\theta \in W : \|h(\theta)\|_2^2 = 0\} = h^{-1}(0) \cap W = \mathcal{M}$, all three functions have the same local zero set. Therefore Theorem 3 applies and yields

$$\mathrm{RLCT}_{\mathcal{M}}(K_1; \pi) = \mathrm{RLCT}_{\mathcal{M}}(K_2; \pi).$$

This proves Corollary 4. □

### C.2. Sharpness under flatter local geometry

The preceding corollary shows that quadratic local geometry is stable. The next lemma identifies the exact change in RLCT under integer powers.

**Lemma 23 (Power scaling of the local zeta function).** *Let $K : U \to [0, \infty)$ be real analytic with local RLCT pair*

$$\mathrm{RLCT}_{\mathcal{M}}(K; \pi) = (\lambda_K, m_K), \qquad \mathcal{M} = \{\theta \in U : K(\theta) = 0\}.$$

*Fix an integer $q \geq 1$. Then $K^q$ is real analytic, has the same zero set $\mathcal{M}$, and*

$$\mathrm{RLCT}_{\mathcal{M}}(K^q; \pi) = \left(\frac{\lambda_K}{q}, m_K\right). \tag{30}$$

*Proof.* For $\Re z > 0$, the local zeta function of $K^q$ is

$$\zeta_{K^q,U}(z) = \int_U K(\theta)^{qz} \pi(\theta)\, d\theta = \zeta_{K,U}(qz).$$

Since both sides are meromorphic continuations of the same holomorphic function on the half-plane $\Re z > 0$, the identity extends meromorphically to all $z \in \mathbb{C}$.

Now let $-\lambda_K$ be the largest pole of $\zeta_{K,U}$, with multiplicity $m_K$. Then the poles of $\zeta_{K^q,U}(z) = \zeta_{K,U}(qz)$ occur exactly at the scaled locations $z = \xi/q$, where $\xi$ ranges over poles of $\zeta_{K,U}$, with the same multiplicities. In particular, the largest pole of $\zeta_{K^q,U}$ is at

$$-\frac{\lambda_K}{q},$$

and its multiplicity is $m_K$. This proves (30). □

We may now prove the sharpness proposition from the main text.

*Proof of Proposition 5.* Assume that $K_1, K_2 : U \to [0, \infty)$ are real analytic with common zero set $\mathcal{M}$, and that

$$K_2 \asymp_{\mathcal{M}} K_1^q \qquad \text{for some integer } q \geq 2.$$

Write

$$\mathrm{RLCT}_{\mathcal{M}}(K_1; \pi) = (\lambda_1, m_1).$$

By Lemma 23,

$$\mathrm{RLCT}_{\mathcal{M}}(K_1^q; \pi) = \left(\frac{\lambda_1}{q}, m_1\right).$$

Since $K_2 \asymp_{\mathcal{M}} K_1^q$ and the two functions have the same local zero set $\mathcal{M}$, Theorem 3 gives

$$\mathrm{RLCT}_{\mathcal{M}}(K_2; \pi) = \mathrm{RLCT}_{\mathcal{M}}(K_1^q; \pi).$$

Therefore,

$$\mathrm{RLCT}_{\mathcal{M}}(K_2; \pi) = \left(\frac{\lambda_1}{q}, m_1\right).$$

If we write

$$\mathrm{RLCT}_{\mathcal{M}}(K_2; \pi) = (\lambda_2, m_2),$$

then

$$\lambda_2 = \frac{\lambda_1}{q}, \qquad m_2 = m_1,$$

which is exactly the claim of Proposition 5. □

**Remark 24 (Why quadratic geometry is the boundary).** *The proofs above show that quadratic local behavior is not merely sufficient for universality; it is the exact threshold at which the learning coefficient remains unchanged. Once the excess-loss germ is flattened to a higher-order power $q \geq 2$, the leading singular complexity changes from $\lambda$ to $\lambda/q$. This is the sense in which Corollary 4 is sharp.*

## D. Gaussian Matrix Identities and Local Norm Equivalences

This appendix proves the matrix identities used in Section 4. The Gaussian negative log-loss is standard, while the Hyvärinen loss is the score-matching criterion introduced by Hyvärinen (2005); see also Jewson and Rossell (2022) for its use in generalized-Bayes updating. Throughout this appendix, $U \subset \Theta$ denotes a fixed compact neighborhood such that

$$\Sigma(U) \subset \mathbb{S}^p_{++}, \qquad \Sigma_0 \in \Sigma(U),$$

and we choose constants $0 < \underline{\kappa} \leq \overline{\kappa} < \infty$ satisfying

$$\underline{\kappa} I_p \preceq \Sigma(\theta) \preceq \overline{\kappa} I_p, \qquad \theta \in U. \tag{31}$$

Since $\Sigma_0 = \Sigma(\theta^\star)$ for some $\theta^\star \in U$, the same bounds hold for $\Sigma_0$ as well:

$$\underline{\kappa} I_p \preceq \Sigma_0 \preceq \overline{\kappa} I_p.$$

### D.1. Exact excess-loss formulas

We begin with the two exact population excess-loss identities stated in Proposition 6.

**Lemma 25 (Closed form of the Gaussian excess log-loss).** *Assume $P_0 = N_p(0, \Sigma_0)$. Then, for every $\theta \in U$,*

$$R_{\ell_{\log}}(\theta) = \frac{1}{2}\Big( \log\det \Sigma(\theta) + \mathrm{tr}\big(\Sigma(\theta)^{-1}\Sigma_0\big)\Big), \tag{32}$$

*and*

$$K_{\log}(\theta) = \frac{1}{2}\Big[ \mathrm{tr}\big(\Sigma(\theta)^{-1}\Sigma_0\big) - \log\det\big(\Sigma(\theta)^{-1}\Sigma_0\big) - p\Big]. \tag{33}$$

*In particular, $K_{\log}(\theta) \geq 0$, with equality if and only if $\Sigma(\theta) = \Sigma_0$.*

*Proof.* Recall that

$$\ell_{\log}(x, \theta) = \frac{1}{2}\Big( \log\det \Sigma(\theta) + x^\top \Sigma(\theta)^{-1} x\Big).$$

Since $X \sim N_p(0, \Sigma_0)$, we have $\mathbb{E}_{P_0}[XX^\top] = \Sigma_0$, and therefore

$$\mathbb{E}_{P_0}\big[X^\top \Sigma(\theta)^{-1} X\big] = \mathrm{tr}\big(\Sigma(\theta)^{-1}\mathbb{E}_{P_0}[XX^\top]\big) = \mathrm{tr}\big(\Sigma(\theta)^{-1}\Sigma_0\big).$$

Substituting into the definition of $R_{\ell_{\log}}$ yields (32).

Now let $\theta^\star \in U$ satisfy $\Sigma(\theta^\star) = \Sigma_0$. Then

$$R_{\ell_{\log}}(\theta^\star) = \frac{1}{2}\Big( \log\det \Sigma_0 + \mathrm{tr}(I_p)\Big) = \frac{1}{2}\big( \log\det \Sigma_0 + p\big).$$

Hence

$$\begin{aligned} K_{\log}(\theta) &= R_{\ell_{\log}}(\theta) - R_{\ell_{\log}}(\theta^\star) \\ &= \frac{1}{2}\Big( \log\det \Sigma(\theta) - \log\det \Sigma_0 + \mathrm{tr}\big(\Sigma(\theta)^{-1}\Sigma_0\big) - p\Big). \end{aligned}$$

Since

$$\log\det\big(\Sigma(\theta)^{-1}\Sigma_0\big) = \log\det \Sigma_0 - \log\det \Sigma(\theta),$$

this is exactly (33).

To prove nonnegativity, define

$$B(\theta) := \Sigma_0^{1/2}\Sigma(\theta)^{-1}\Sigma_0^{1/2}.$$

Then $B(\theta) \in \mathbb{S}^p_{++}$, and

$$\mathrm{tr}\big(\Sigma(\theta)^{-1}\Sigma_0\big) = \mathrm{tr}(B(\theta)), \qquad \log\det\big(\Sigma(\theta)^{-1}\Sigma_0\big) = \log\det B(\theta).$$

If $\mu_1(\theta), \ldots, \mu_p(\theta)$ are the eigenvalues of $B(\theta)$, then

$$K_{\log}(\theta) = \frac{1}{2}\sum_{i=1}^{p} \big(\mu_i(\theta) - \log\mu_i(\theta) - 1\big).$$

Since the scalar function $t \mapsto t - \log t - 1$ is nonnegative on $(0, \infty)$ and vanishes only at $t = 1$, it follows that $K_{\log}(\theta) \geq 0$, with equality if and only if $\mu_i(\theta) = 1$ for all $i$, that is, $B(\theta) = I_p$. Equivalently, $\Sigma(\theta) = \Sigma_0$. □

**Lemma 26 (Closed form of the Hyvärinen excess loss).** *Assume $P_0 = N_p(0, \Sigma_0)$. Then, for every $\theta \in U$,*

$$R_{\ell_{\mathrm{H}}}(\theta) = \frac{1}{2}\,\mathrm{tr}\big(\Sigma(\theta)^{-2}\Sigma_0\big) - \mathrm{tr}\big(\Sigma(\theta)^{-1}\big), \tag{34}$$

*and*

$$K_{\mathrm{H}}(\theta) = \frac{1}{2}\operatorname{tr}\Big[\Sigma_0\big(\Sigma(\theta)^{-1} - \Sigma_0^{-1}\big)^2\Big]. \tag{35}$$

*Equivalently,*

$$K_{\mathrm{H}}(\theta) = \frac{1}{2}\left\|\big(\Sigma(\theta)^{-1} - \Sigma_0^{-1}\big)\Sigma_0^{1/2}\right\|_{\mathrm{F}}^2. \tag{36}$$

*In particular,* $K_{\mathrm{H}}(\theta) \geq 0$*, with equality if and only if* $\Sigma(\theta) = \Sigma_0$*.*

*Proof.* Recall that

$$\ell_{\mathrm{H}}(x,\theta) = \frac{1}{2}x^\top \Sigma(\theta)^{-2}x - \operatorname{tr}\big(\Sigma(\theta)^{-1}\big).$$

As above,

$$\mathbb{E}_{P_0}\big[X^\top \Sigma(\theta)^{-2}X\big] = \operatorname{tr}\big(\Sigma(\theta)^{-2}\Sigma_0\big),$$

which proves (34).

Now fix $\theta^\star \in U$ such that $\Sigma(\theta^\star) = \Sigma_0$. Then

$$R_{\ell_{\mathrm{H}}}(\theta^\star) = \frac{1}{2}\operatorname{tr}\big(\Sigma_0^{-1}\big) - \operatorname{tr}\big(\Sigma_0^{-1}\big) = -\frac{1}{2}\operatorname{tr}\big(\Sigma_0^{-1}\big).$$

Therefore

$$\begin{aligned} K_{\mathrm{H}}(\theta) &= R_{\ell_{\mathrm{H}}}(\theta) - R_{\ell_{\mathrm{H}}}(\theta^\star) \\ &= \frac{1}{2}\operatorname{tr}\big(\Sigma(\theta)^{-2}\Sigma_0\big) - \operatorname{tr}\big(\Sigma(\theta)^{-1}\big) + \frac{1}{2}\operatorname{tr}\big(\Sigma_0^{-1}\big). \end{aligned}$$

Set

$$\Delta(\theta) := \Sigma(\theta)^{-1} - \Sigma_0^{-1}.$$

Since both $\Sigma(\theta)^{-1}$ and $\Sigma_0^{-1}$ are symmetric, $\Delta(\theta)$ is symmetric. Expanding the square gives

$$\begin{aligned} \operatorname{tr}\big(\Sigma_0\Delta(\theta)^2\big) &= \operatorname{tr}\big(\Sigma_0\Sigma(\theta)^{-2}\big) - \operatorname{tr}\big(\Sigma_0\Sigma(\theta)^{-1}\Sigma_0^{-1}\big) - \operatorname{tr}\big(\Sigma_0\Sigma_0^{-1}\Sigma(\theta)^{-1}\big) + \operatorname{tr}\big(\Sigma_0\Sigma_0^{-2}\big) \\ &= \operatorname{tr}\big(\Sigma(\theta)^{-2}\Sigma_0\big) - 2\operatorname{tr}\big(\Sigma(\theta)^{-1}\big) + \operatorname{tr}\big(\Sigma_0^{-1}\big), \end{aligned}$$

where we used cyclicity of the trace in the middle terms. Multiplying by $1/2$ yields (35).

For (36), note that

$$\operatorname{tr}\big(\Sigma_0\Delta(\theta)^2\big) = \operatorname{tr}\big(\Sigma_0^{1/2}\Delta(\theta)^2\Sigma_0^{1/2}\big) = \operatorname{tr}\Big(\big(\Delta(\theta)\Sigma_0^{1/2}\big)^\top\big(\Delta(\theta)\Sigma_0^{1/2}\big)\Big),$$

because $\Delta(\theta)$ is symmetric. Hence

$$\operatorname{tr}\big(\Sigma_0\Delta(\theta)^2\big) = \left\|\Delta(\theta)\Sigma_0^{1/2}\right\|_{\mathrm{F}}^2.$$

The nonnegativity and zero-set claim now follow immediately from (36): the expression is a squared Frobenius norm and vanishes if and only if $\Delta(\theta) = 0$, that is, $\Sigma(\theta) = \Sigma_0$. □

### D.2. Auxiliary norm comparison lemmas

We next prove the norm-comparison lemmas needed to convert the exact identities above into the local quadratic equivalences from Theorem 7.

**Lemma 27 (Spectral sandwich for $t - \log t - 1$).** *Let* $0 < a < b < \infty$*. Then there exist constants* $0 < c_{a,b} \leq C_{a,b} < \infty$ *such that*

$$c_{a,b}(t-1)^2 \leq t - \log t - 1 \leq C_{a,b}(t-1)^2, \qquad t \in [a,b]. \tag{37}$$

*Proof.* Define

$$\phi(t) := t - \log t - 1, \qquad t > 0.$$

For $t \neq 1$, set

$$\psi(t) := \frac{\phi(t)}{(t-1)^2}.$$

A second-order Taylor expansion of $\phi$ around $t = 1$ gives

$$\phi(t) = \frac{1}{2}(t-1)^2 + o\big((t-1)^2\big),$$

so $\psi$ extends continuously to $t = 1$ by setting $\psi(1) = 1/2$. Because $\phi(t) > 0$ for all $t \neq 1$, the extension of $\psi$ is continuous and strictly positive on the compact interval $[a, b]$. Hence

$$0 < c_{a,b} := \min_{t\in[a,b]} \psi(t) \leq \max_{t\in[a,b]} \psi(t) =: C_{a,b} < \infty.$$

Multiplying through by $(t-1)^2$ yields (37). □

**Lemma 28 (Frobenius norm under fixed SPD congruence).** *Let $S \in \mathbb{S}^p_{++}$ satisfy*

$$mI_p \preceq S \preceq MI_p$$

*for some $0 < m \leq M < \infty$. Then, for every $A \in \mathbb{R}^{p\times p}$,*

$$m \, \|A\|_{\mathrm{F}} \leq \left\|S^{1/2}AS^{1/2}\right\|_{\mathrm{F}} \leq M \, \|A\|_{\mathrm{F}} . \tag{38}$$

*Proof.* The upper bound follows from submultiplicativity of the Frobenius norm with the operator norm:

$$\left\|S^{1/2}AS^{1/2}\right\|_{\mathrm{F}} \leq \left\|S^{1/2}\right\|^2_{\mathrm{op}} \|A\|_{\mathrm{F}} \leq M \, \|A\|_{\mathrm{F}} .$$

For the lower bound, write

$$A = S^{-1/2}\big(S^{1/2}AS^{1/2}\big)S^{-1/2}.$$

Applying the same estimate to the right-hand side gives

$$\|A\|_{\mathrm{F}} \leq \left\|S^{-1/2}\right\|^2_{\mathrm{op}} \left\|S^{1/2}AS^{1/2}\right\|_{\mathrm{F}} \leq m^{-1} \left\|S^{1/2}AS^{1/2}\right\|_{\mathrm{F}},$$

which is equivalent to the lower bound in (38). □

**Lemma 29 (Weighted Frobenius norm for symmetric matrices).** *Let $M \in \mathbb{S}^p_{++}$ satisfy*

$$mI_p \preceq M \preceq M_+ I_p$$

*for some $0 < m \leq M_+ < \infty$. If $A \in \mathbb{R}^{p\times p}$ is symmetric, then*

$$m \, \|A\|^2_{\mathrm{F}} \leq \mathrm{tr}\big(MA^2\big) \leq M_+ \, \|A\|^2_{\mathrm{F}} . \tag{39}$$

*Proof.* Since $A$ is symmetric, $A^2$ is positive semidefinite. Therefore

$$m \, \mathrm{tr}(A^2) \leq \mathrm{tr}(MA^2) \leq M_+ \, \mathrm{tr}(A^2).$$

Because $A$ is symmetric,

$$\mathrm{tr}(A^2) = \mathrm{tr}(A^\top A) = \|A\|^2_{\mathrm{F}},$$

and (39) follows. □

**Lemma 30 (Bi-Lipschitz behavior of matrix inversion on bounded SPD sets).** *Under* (31)*, for every* $\theta \in U$*,*

$$\overline{\kappa}^{-2} \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}} \le \left\|\Sigma(\theta)^{-1} - \Sigma_0^{-1}\right\|_{\mathrm{F}} \le \underline{\kappa}^{-2} \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}. \tag{40}$$

*Proof.* The resolvent identity gives

$$\Sigma(\theta)^{-1} - \Sigma_0^{-1} = \Sigma(\theta)^{-1}\big(\Sigma_0 - \Sigma(\theta)\big)\Sigma_0^{-1}.$$

Hence

$$\left\|\Sigma(\theta)^{-1} - \Sigma_0^{-1}\right\|_{\mathrm{F}} \le \left\|\Sigma(\theta)^{-1}\right\|_{\mathrm{op}} \left\|\Sigma_0^{-1}\right\|_{\mathrm{op}} \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}} \le \underline{\kappa}^{-2} \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}},$$

which is the upper bound.

For the lower bound, rearrange the same identity as

$$\Sigma(\theta) - \Sigma_0 = \Sigma(\theta)\big(\Sigma_0^{-1} - \Sigma(\theta)^{-1}\big)\Sigma_0.$$

Therefore

$$\|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}} \le \|\Sigma(\theta)\|_{\mathrm{op}} \|\Sigma_0\|_{\mathrm{op}} \left\|\Sigma(\theta)^{-1} - \Sigma_0^{-1}\right\|_{\mathrm{F}} \le \overline{\kappa}^2 \left\|\Sigma(\theta)^{-1} - \Sigma_0^{-1}\right\|_{\mathrm{F}},$$

which is equivalent to the lower bound in (40). □

### D.3. Proof of local quadratic equivalence

We now assemble the preceding lemmas to prove Theorem 7.

*Proof of Theorem 7.* Fix $\theta \in U$, and define

$$\Delta(\theta) := \Sigma(\theta)^{-1} - \Sigma_0^{-1}.$$

**Step 1: the Hyvärinen excess loss.** By Lemma 26,

$$K_{\mathrm{H}}(\theta) = \frac{1}{2} \operatorname{tr}\big(\Sigma_0 \Delta(\theta)^2\big).$$

Since $\Delta(\theta)$ is symmetric, Lemma 29 with $M = \Sigma_0$ yields

$$\frac{1}{2}\underline{\kappa} \|\Delta(\theta)\|_{\mathrm{F}}^2 \le K_{\mathrm{H}}(\theta) \le \frac{1}{2}\overline{\kappa} \|\Delta(\theta)\|_{\mathrm{F}}^2.$$

Applying Lemma 30 gives

$$\frac{1}{2}\underline{\kappa}\,\overline{\kappa}^{-4} \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2 \le K_{\mathrm{H}}(\theta) \le \frac{1}{2}\overline{\kappa}\,\underline{\kappa}^{-4} \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2.$$

Thus $K_{\mathrm{H}}(\theta) \asymp \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2$ uniformly on $U$.

**Step 2: the Gaussian excess log-loss.** Define

$$B(\theta) := \Sigma_0^{1/2} \Sigma(\theta)^{-1} \Sigma_0^{1/2}.$$

Then $B(\theta) \in \mathbb{S}_{++}^p$, and by Lemma 25,

$$K_{\log}(\theta) = \frac{1}{2}\Big(\operatorname{tr}(B(\theta)) - \log\det B(\theta) - p\Big).$$

Let $\mu_1(\theta), \dots, \mu_p(\theta)$ denote the eigenvalues of $B(\theta)$. Then

$$K_{\log}(\theta) = \frac{1}{2} \sum_{i=1}^{p} \phi(\mu_i(\theta)), \qquad \phi(t) := t - \log t - 1.$$

We claim that all eigenvalues of $B(\theta)$ lie in a fixed compact interval $[a, b] \subset (0, \infty)$, uniformly in $\theta \in U$. Indeed, using (31),

$$\lambda_{\min}\big(B(\theta)\big) \geq \lambda_{\min}(\Sigma_0)\lambda_{\min}\big(\Sigma(\theta)^{-1}\big) \geq \underline{\kappa}\,\overline{\kappa}^{-1} =: a,$$

and similarly

$$\lambda_{\max}\big(B(\theta)\big) \leq \lambda_{\max}(\Sigma_0)\lambda_{\max}\big(\Sigma(\theta)^{-1}\big) \leq \overline{\kappa}\,\underline{\kappa}^{-1} =: b.$$

Hence $\mu_i(\theta) \in [a, b]$ for every $i$ and every $\theta \in U$.

Applying Lemma 27 on $[a, b]$, we obtain constants $0 < c_{a,b} \leq C_{a,b} < \infty$ such that

$$\frac{c_{a,b}}{2} \sum_{i=1}^{p} \big(\mu_i(\theta) - 1\big)^2 \leq K_{\log}(\theta) \leq \frac{C_{a,b}}{2} \sum_{i=1}^{p} \big(\mu_i(\theta) - 1\big)^2.$$

Since $B(\theta)$ is symmetric,

$$\sum_{i=1}^{p} \big(\mu_i(\theta) - 1\big)^2 = \|B(\theta) - I_p\|_{\mathrm{F}}^2 .$$

Moreover,

$$B(\theta) - I_p = \Sigma_0^{1/2}\big(\Sigma(\theta)^{-1} - \Sigma_0^{-1}\big)\Sigma_0^{1/2} = \Sigma_0^{1/2}\Delta(\theta)\Sigma_0^{1/2}.$$

Therefore Lemma 28 with $S = \Sigma_0$ gives

$$\underline{\kappa}\, \|\Delta(\theta)\|_{\mathrm{F}} \leq \|B(\theta) - I_p\|_{\mathrm{F}} \leq \overline{\kappa}\, \|\Delta(\theta)\|_{\mathrm{F}} .$$

Combining the previous displays,

$$\frac{c_{a,b}}{2}\underline{\kappa}^2 \|\Delta(\theta)\|_{\mathrm{F}}^2 \leq K_{\log}(\theta) \leq \frac{C_{a,b}}{2}\overline{\kappa}^2 \|\Delta(\theta)\|_{\mathrm{F}}^2 .$$

A final application of Lemma 30 yields

$$\frac{c_{a,b}}{2}\underline{\kappa}^2\overline{\kappa}^{-4} \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2 \leq K_{\log}(\theta) \leq \frac{C_{a,b}}{2}\overline{\kappa}^2\underline{\kappa}^{-4} \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2 .$$

Thus $K_{\log}(\theta) \asymp \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2$ uniformly on $U$.

**Step 3: conclusion.** We have proved two-sided bounds of the form

$$c_1 \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2 \leq K_{\log}(\theta) \leq C_1 \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2$$

and

$$c_2 \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2 \leq K_{\mathrm{H}}(\theta) \leq C_2 \|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2$$

for all $\theta \in U$. In particular,

$$K_{\log} \asymp_{\mathcal{M}(\Sigma_0)} K_{\mathrm{H}},$$

which is exactly the content of Theorem 7. □

## E. Universality in Analytic Covariance Families

This appendix proves Corollary 8. The point is to make explicit the covariance-coordinate structure underlying the argument. Once the target covariance $\Sigma_0$ is fixed, both the Gaussian excess log-loss and the Hyvärinen excess loss factor through the same analytic map

$$h(\theta) = \mathrm{vech}\big(\Sigma(\theta) - \Sigma_0\big),$$

which cuts out the covariance fiber. The remaining task is then to show that the induced germs in the $h$-coordinates are analytic and nondegenerate quadratic at the origin, so that Corollary 4 applies directly.

### E.1. Covariance coordinates and exact factorization

Let $\Sigma : \Theta \to \mathbb{S}^p_{++}$ be a real-analytic covariance map and fix $\Sigma_0$ in its image. Recall the covariance fiber

$$\mathcal{M}(\Sigma_0) = \{\theta \in \Theta : \Sigma(\theta) = \Sigma_0\}.$$

We first isolate the coordinate map that describes this fiber.

**Lemma 31 (Analytic covariance coordinates).** *Define*

$$h : \Theta \to \mathbb{R}^{p(p+1)/2}, \qquad h(\theta) := \mathrm{vech}\big(\Sigma(\theta) - \Sigma_0\big). \tag{41}$$

*Then $h$ is real analytic and*

$$h^{-1}(0) = \mathcal{M}(\Sigma_0). \tag{42}$$

*Moreover, let* $\mathrm{mat} : \mathbb{R}^{p(p+1)/2} \to \mathbb{R}^{p\times p}$ *denote the linear inverse of* vech *on symmetric matrices. There exists $\delta > 0$ such that*

$$\Sigma_0 + \mathrm{mat}(u) \in \mathbb{S}^p_{++}, \qquad \|u\|_2 < \delta,$$

*and the functions*

$$g_{\log}(u) := \frac{1}{2}\Big[\,\mathrm{tr}\big((\Sigma_0 + \mathrm{mat}(u))^{-1}\Sigma_0\big) - \log\det\big((\Sigma_0 + \mathrm{mat}(u))^{-1}\Sigma_0\big) - p\Big], \tag{43}$$

$$g_{\mathrm{H}}(u) := \frac{1}{2}\,\mathrm{tr}\Big[\Sigma_0\big((\Sigma_0 + \mathrm{mat}(u))^{-1} - \Sigma_0^{-1}\big)^2\Big] \tag{44}$$

*are real analytic on the Euclidean ball $\{u : \|u\|_2 < \delta\}$. Finally, for every $\theta$ such that $\|h(\theta)\|_2 < \delta$,*

$$K_{\log}(\theta) = g_{\log}(h(\theta)), \qquad K_{\mathrm{H}}(\theta) = g_{\mathrm{H}}(h(\theta)). \tag{45}$$

*Proof.* Since $\Sigma$ is real analytic as a matrix-valued map, each coordinate function $\theta \mapsto \Sigma_{ij}(\theta)$ is real analytic. Because vech is linear, the components of $h$ are real analytic as well, proving analyticity of (41).

Now $h(\theta) = 0$ if and only if $\mathrm{vech}(\Sigma(\theta) - \Sigma_0) = 0$, which for a symmetric matrix is equivalent to $\Sigma(\theta) - \Sigma_0 = 0$. Hence (42) holds.

To prove the existence of $\delta$, note that $\mathbb{S}^p_{++}$ is an open subset of the vector space of symmetric $p \times p$ matrices, and $\Sigma_0 \in \mathbb{S}^p_{++}$. Therefore there exists $\varepsilon > 0$ such that every symmetric matrix $S$ satisfying

$$\|S - \Sigma_0\|_{\mathrm{F}} < \varepsilon$$

also lies in $\mathbb{S}^p_{++}$. By Lemma 32 below, for every symmetric matrix $A$,

$$\|A\|_{\mathrm{F}} \le \sqrt{2}\,\|\mathrm{vech}(A)\|_2\,.$$

Hence if $\|u\|_2 < \varepsilon/\sqrt{2}$, then

$$\|\mathrm{mat}(u)\|_{\mathrm{F}} < \varepsilon,$$

so $\Sigma_0 + \mathrm{mat}(u) \in \mathbb{S}^p_{++}$. We may therefore take

$$\delta := \varepsilon/\sqrt{2}.$$

The maps

$$u \mapsto \Sigma_0 + \mathrm{mat}(u), \qquad S \mapsto S^{-1}, \qquad S \mapsto \log\det S,$$

are real analytic on their respective domains, and the formulas (43)–(44) are obtained from these operations by algebraic composition. Thus $g_{\log}$ and $g_{\mathrm{H}}$ are real analytic on $\|u\|_2 < \delta$.

Finally, for every $\theta$ with $\|h(\theta)\|_2 < \delta$,

$$\Sigma_0 + \mathrm{mat}(h(\theta)) = \Sigma_0 + \Sigma(\theta) - \Sigma_0 = \Sigma(\theta).$$

Substituting this into (43) and (44), and using the exact formulas from Lemmas 25 and 26, yields (45). □

The next lemma compares Euclidean norm in the half-vectorized coordinates with Frobenius norm on symmetric matrices.

**Lemma 32 (Half-vectorization versus Frobenius norm).** *For every symmetric matrix $A = (a_{ij}) \in \mathbb{R}^{p\times p}$,*

$$\|\mathrm{vech}(A)\|_2^2 \le \|A\|_{\mathrm{F}}^2 \le 2\,\|\mathrm{vech}(A)\|_2^2\,. \tag{46}$$

*Equivalently, for every $u \in \mathbb{R}^{p(p+1)/2}$,*

$$\|u\|_2^2 \le \|\mathrm{mat}(u)\|_{\mathrm{F}}^2 \le 2\,\|u\|_2^2\,. \tag{47}$$

*Proof.* For a symmetric matrix $A$,

$$\|A\|_{\mathrm{F}}^2 = \sum_{i=1}^{p} a_{ii}^2 + 2\sum_{1\le j<i\le p} a_{ij}^2,$$

whereas

$$\|\mathrm{vech}(A)\|_2^2 = \sum_{i=1}^{p} a_{ii}^2 + \sum_{1\le j<i\le p} a_{ij}^2.$$

The inequalities in (46) follow immediately by comparing the two expressions term by term. Since mat is the inverse of vech on symmetric matrices, (47) is equivalent to (46). □

### E.2. Nondegenerate quadratic germs in covariance coordinates

We now show that the germs induced by (43) and (44) are genuinely quadratic at the origin.

**Lemma 33 (Nondegenerate quadratic germs in covariance coordinates).** *The analytic germs $g_{\log}$ and $g_{\mathrm{H}}$ from (43)–(44) satisfy*

$$g_j(0) = 0, \qquad \nabla g_j(0) = 0, \qquad \nabla^2 g_j(0) \succ 0, \qquad j \in \{\log, \mathrm{H}\}.$$

*Equivalently, both are nondegenerate quadratic germs at the origin.*

*Proof.* We give the argument for a generic $g_j$, with $j \in \{\log, \mathrm{H}\}$.

**Step 1: local quadratic sandwich in the $u$-coordinates.** Consider the local covariance family

$$\Sigma_u := \Sigma_0 + \mathrm{mat}(u), \qquad \|u\|_2 < \delta.$$

For $\delta$ sufficiently small, the set $\{\Sigma_u : \|u\|_2 < \delta\}$ is contained in a compact eigenvalue-bounded neighborhood of $\Sigma_0$ inside $\mathbb{S}^p_{++}$. Therefore Theorem 7 applies to this local family and yields constants $0 < c_j \le C_j < \infty$ such that

$$c_j \|\Sigma_u - \Sigma_0\|_{\mathrm{F}}^2 \le g_j(u) \le C_j \|\Sigma_u - \Sigma_0\|_{\mathrm{F}}^2, \qquad \|u\|_2 < \delta.$$

Since $\Sigma_u - \Sigma_0 = \mathrm{mat}(u)$, Lemma 32 implies

$$c_j \|u\|_2^2 \le g_j(u) \le 2C_j \|u\|_2^2, \qquad \|u\|_2 < \delta.$$

Thus $g_j$ is locally equivalent to $\|u\|_2^2$.

**Step 2: value and gradient at the origin.** By construction,

$$g_j(0) = 0,$$

because $u = 0$ corresponds to the target covariance $\Sigma_0$, at which both excess losses vanish. Moreover, the inequality $g_j(u) \ge 0$ for $\|u\|_2 < \delta$ shows that $u = 0$ is a local minimizer. Since $g_j$ is differentiable, this implies

$$\nabla g_j(0) = 0.$$

**Step 3: positive definiteness of the Hessian.** Fix $v \in \mathbb{R}^{p(p+1)/2}$, $v \neq 0$. By analyticity of $g_j$ and the identities $g_j(0) = 0$, $\nabla g_j(0) = 0$, Taylor expansion gives

$$g_j(tv) = \frac{t^2}{2} v^\top \nabla^2 g_j(0)\, v + o(t^2) \qquad \text{as } t \to 0.$$

On the other hand, the lower quadratic bound from Step 1 implies

$$g_j(tv) \ge c_j t^2 \|v\|_2^2 \qquad \text{for all sufficiently small } t.$$

Dividing by $t^2$ and letting $t \to 0$ yields

$$\frac{1}{2} v^\top \nabla^2 g_j(0)\, v \ge c_j \|v\|_2^2 > 0.$$

Since this holds for every nonzero $v$, the Hessian $\nabla^2 g_j(0)$ is positive definite. □

### E.3. Proof of universality in analytic covariance families

We can now complete the proof of Corollary 8.

*Proof of Corollary 8.* Fix the analytic covariance map $\Sigma : \Theta \to \mathbb{S}^p_{++}$ and target covariance $\Sigma_0$ in its image. By Lemma 31, the map

$$h(\theta) = \mathrm{vech}\big(\Sigma(\theta) - \Sigma_0\big)$$

is real analytic and satisfies

$$h^{-1}(0) = \mathcal{M}(\Sigma_0).$$

Moreover, on a neighborhood of the origin in $\mathbb{R}^{p(p+1)/2}$, there exist real-analytic functions $g_{\log}$ and $g_{\mathrm{H}}$ such that

$$K_{\log}(\theta) = g_{\log}(h(\theta)), \qquad K_{\mathrm{H}}(\theta) = g_{\mathrm{H}}(h(\theta)).$$

By Lemma 33,

$$g_j(0) = 0, \qquad \nabla g_j(0) = 0, \qquad \nabla^2 g_j(0) \succ 0, \qquad j \in \{\log, \mathrm{H}\}.$$

Therefore the hypotheses of Corollary 4 are satisfied with common analytic map $h$, minimizer set

$$\mathcal{M} = h^{-1}(0) = \mathcal{M}(\Sigma_0),$$

and analytic germs $g_{\log}$, $g_{\mathrm{H}}$. Hence

$$\mathrm{RLCT}_{\mathcal{M}(\Sigma_0)}(K_{\log}; \pi) = \mathrm{RLCT}_{\mathcal{M}(\Sigma_0)}(K_{\mathrm{H}}; \pi).$$

This is exactly the statement of Corollary 8. □

**Remark 34 (Alternative route via direct local comparison).** *The preceding proof is structured through Corollary 4: both losses factor through the same covariance-coordinate map and have nondegenerate quadratic germs in those coordinates. One could also obtain Corollary 8 directly from Theorem 7 and Theorem 3. We prefer the present formulation because it makes the geometry explicit: the common analytic minimizer map is simply the covariance fiber map $\theta \mapsto \Sigma(\theta) - \Sigma_0$.*

## F. Factor-Analysis Transfer Details

This appendix makes explicit which factor-analysis results are imported from the recent singular-learning-theory analysis of Drton et al. (2025) and which part is new in the present paper. The new statement is the transfer theorem Theorem 9: the local RLCT pair for the Hyvärinen generalized posterior equals that of the ordinary Gaussian posterior at every factor-analysis covariance fiber. No new resolution-of-singularities computation for factor analysis is carried out here. Once Theorem 9 is available, all Bayes-side formulas and bounds from Drton et al. (2025) carry over verbatim.

### F.1. Imported Bayes-side inputs

Recall the factor-analysis covariance model

$$\Sigma(\psi, \Lambda) = \Psi + \Lambda\Lambda^\top, \qquad \Psi = \mathrm{diag}(\psi_1, \dots, \psi_p), \qquad (\psi, \Lambda) \in \Theta^{\mathrm{FA}}_{p,k}.$$

For a covariance matrix $\Sigma_0 \in \mathcal{C}^{\mathrm{FA}}_{p,k}$, write

$$r(\Sigma_0) := \min\Big\{ \mathrm{rank}(A) : \Sigma_0 = \mathrm{diag}(\psi) + AA^\top \text{ for some } \psi \in (0,\infty)^p,\ A \in \mathbb{R}^{p\times k} \Big\},$$

and set

$$d_r := (r+1)p - \frac{r(r-1)}{2}.$$

The Bayes-side learning coefficient and multiplicity are denoted by

$$(\lambda^{\log}_{p,k}(\Sigma_0), m^{\log}_{p,k}(\Sigma_0)) := \mathrm{RLCT}_{\mathcal{M}_{p,k}(\Sigma_0)}(K_{\log}; \pi).$$

The following proposition is not new; it is a reformulation of the exact results proved in Drton et al. (2025) in the notation of the present paper.

**Proposition 35 (Imported Bayes-side factor-analysis results).** *Let $\Sigma_0 \in \mathcal{C}^{\mathrm{FA}}_{p,k}$.*

1. ***General upper bound.*** *If $d_{r(\Sigma_0)} \le p(p+1)/2$, then*

$$\lambda^{\log}_{p,k}(\Sigma_0) \le \frac{p(k+2) + r(\Sigma_0)\big(p-k+1\big)}{4}. \tag{48}$$

2. ***Diagonal case*** $(r = 0)$**.** *If $\Sigma_0$ is diagonal, then*

$$\lambda^{\log}_{p,k}(\Sigma_0) = \begin{cases} \dfrac{p(k+2)}{4}, & k \le p-1, \\ \dfrac{p^2+p+1}{4}, & k = p, \end{cases} \tag{49}$$

*and*

$$m^{\log}_{p,k}(\Sigma_0) = \begin{cases} 1, & k \le p-2, \ \text{or } k = p, \ \text{or } (p,k) = (1,0), \\ p-1, & k = p-1 > 0. \end{cases} \tag{50}$$

3. ***Generic rank-one case*** $(r = 1)$**.** *If $\Sigma_0$ is chosen generically from the one-factor submodel, then*

$$\lambda^{\log}_{p,k}(\Sigma_0) = \begin{cases} \dfrac{pk+3p-k+1}{4}, & k \le p-2, \\ \dfrac{p(p+1)}{4}, & k \in \{p-1, p\}, \end{cases} \tag{51}$$

*and*

$$m^{\log}_{p,k}(\Sigma_0) = 1. \tag{52}$$

4. ***Generic full-rank case*** $(r = k)$**.** *If $\Sigma_0$ is chosen generically from the $k$-factor model and $d_k \le p(p+1)/2$, then*

$$\lambda^{\log}_{p,k}(\Sigma_0) = \frac{d_k}{2}, \qquad m^{\log}_{p,k}(\Sigma_0) = 1. \tag{53}$$

*Proof.* The upper bound (48) is Theorem 1.1 of Drton et al. (2025). The tightness statements for $r = 0$, generic $r = 1$, and generic $r = k$ are summarized in their Proposition 1.2. The diagonal formulas (49)–(50) are their Theorem 4.6. The generic rank-one formulas (51)–(52) are their Theorem 4.15. The generic full-rank formula (53) is their Lemma 3.4(1), specialized to the generic $k$-factor stratum. □

### F.2. Proof of exact transfer to factor analysis

We can now prove the factor-analysis transfer theorem from the main text.

*Proof of Theorem 9.* The factor-analysis covariance map

$$(\psi, \Lambda) \longmapsto \Psi + \Lambda\Lambda^\top$$

is real analytic on $\Theta^{\mathrm{FA}}_{p,k}$, and its image is contained in $\mathbb{S}^p_{++}$. For a fixed target covariance $\Sigma_0 \in \mathcal{C}^{\mathrm{FA}}_{p,k}$, the associated covariance fiber in the sense of Corollary 8 is precisely

$$\mathcal{M}(\Sigma_0) = \{(\psi, \Lambda) \in \Theta^{\mathrm{FA}}_{p,k} : \Psi + \Lambda\Lambda^\top = \Sigma_0\} = \mathcal{M}_{p,k}(\Sigma_0).$$

Therefore Corollary 8 applies with $\Theta = \Theta^{\mathrm{FA}}_{p,k}$ and yields

$$\mathrm{RLCT}_{\mathcal{M}_{p,k}(\Sigma_0)}(K_{\log}; \pi) = \mathrm{RLCT}_{\mathcal{M}_{p,k}(\Sigma_0)}(K_{\mathrm{H}}; \pi).$$

This is exactly the statement of Theorem 9. □

The next corollary from the main text is now immediate: all imported Bayes-side formulas transfer unchanged to the Hyvärinen posterior.

*Proof of Corollary 10.* By Theorem 9,

$$\mathrm{RLCT}_{\mathcal{M}_{p,k}(\Sigma_0)}(K_{\mathrm{H}};\pi) = \mathrm{RLCT}_{\mathcal{M}_{p,k}(\Sigma_0)}(K_{\log};\pi),$$

so

$$\lambda^{\mathrm{H}}_{p,k}(\Sigma_0) = \lambda^{\log}_{p,k}(\Sigma_0), \qquad m^{\mathrm{H}}_{p,k}(\Sigma_0) = m^{\log}_{p,k}(\Sigma_0).$$

The transferred upper bound and all stated exact values follow by substituting the Bayes-side formulas from Proposition 35. No further argument is required. □

### F.3. The one-factor trichotomy

We now give the details behind Corollary 11. Write $k = 1$ and parameterize

$$\Sigma_0 = \mathrm{diag}(\psi_0) + a_0 a_0^\top, \qquad \psi_0 \in (0,\infty)^p, \quad a_0 \in \mathbb{R}^p.$$

Since the off-diagonal entries of $\Sigma_0$ are

$$(\Sigma_0)_{ij} = a_{0,i} a_{0,j}, \qquad i \neq j,$$

the support pattern of $a_0$ determines the covariance stratum.

**Lemma 36 (Support pattern and covariance stratum).** *Let* $\Sigma_0 = \mathrm{diag}(\psi_0) + a_0 a_0^\top$ *with* $\psi_0 \in (0,\infty)^p$.

1. $\Sigma_0$ *is diagonal if and only if* $a_0$ *has at most one nonzero entry.*
2. $\Sigma_0$ *has precisely one nonzero off-diagonal pair if and only if* $a_0$ *has exactly two nonzero entries.*

*Proof.* For $i \neq j$, the $(i,j)$-entry of $\Sigma_0$ is $a_{0,i}a_{0,j}$. Hence all off-diagonal entries vanish if and only if no two distinct coordinates of $a_0$ are simultaneously nonzero, that is, if and only if $a_0$ has at most one nonzero entry. This proves (1).

If $a_0$ has exactly two nonzero entries, say at indices $i$ and $j$, then the only potentially nonzero off-diagonal entries are $(i,j)$ and $(j,i)$, and these are indeed nonzero. Conversely, if $\Sigma_0$ has precisely one nonzero off-diagonal pair, then there exist exactly two indices $i \neq j$ such that $a_{0,i}a_{0,j} \neq 0$, while $a_{0,r}a_{0,s} = 0$ for every other unordered pair $\{r,s\}$. This forces exactly two coordinates of $a_0$ to be nonzero. Thus (2) holds. □

We can now complete the one-factor transfer proof.

*Proof of Corollary 11.* Consider the three cases from the statement.

**Case 1: $a_0$ has more than two nonzero entries.** At the start of Section 5, Drton et al. (2025) note that the singularities of the one-factor parametrization occur exactly at loading vectors with at most two nonzero entries. Therefore the present case is the generic one-factor stratum. By Lemma 3.4(1) of Drton et al. (2025), specialized to $k = 1$, the ordinary Gaussian learning coefficient and multiplicity satisfy

$$\lambda^{\log}_{p,1}(\Sigma_0) = p, \qquad m^{\log}_{p,1}(\Sigma_0) = 1.$$

Applying Theorem 9 gives the same values for the Hyvärinen posterior.

**Case 2: $a_0$ has exactly two nonzero entries.** By Lemma 36, this is equivalent to $\Sigma_0$ having precisely one nonzero off-diagonal pair. Proposition 5.1 of Drton et al. (2025) then gives

$$\lambda^{\log}_{p,1}(\Sigma_0) = \frac{2p-1}{2}, \qquad m^{\log}_{p,1}(\Sigma_0) = 1.$$

Another application of Theorem 9 transfers these values to the Hyvärinen posterior.

**Case 3: $a_0$ has at most one nonzero entry.** By Lemma 36, $\Sigma_0$ is diagonal. Proposition 4.2 of Drton et al. (2025) gives

$$\lambda^{\log}_{p,1}(\Sigma_0) = \frac{3p}{4}, \qquad m^{\log}_{p,1}(\Sigma_0) = 1.$$

Applying Theorem 9 once more yields

$$\lambda^{\mathrm{H}}_{p,1}(\Sigma_0) = \frac{3p}{4}, \qquad m^{\mathrm{H}}_{p,1}(\Sigma_0) = 1.$$

These three cases are exhaustive and mutually exclusive, so Corollary 11 follows. □

**Remark 37 (What is new and what is imported).** *The formulas in Proposition 35 and the one-factor values used above are not rederived here. All such calculations are imported from Drton et al. (2025). The new content of the present paper is the exact transport of those Bayes-side RLCT statements to the Hyvärinen generalized posterior via Theorem 9.*

## G. Numerical Methods

This appendix gives the exact numerical protocol used for the experiments in Section 6 and Appendix H. The numerical study is designed to validate theorem-level statements rather than to benchmark a new inference algorithm. Accordingly, the experiments use synthetic data from the same Gaussian factor-analysis families that appear in the theory, and every plotted quantity is either computed exactly from closed-form formulas or estimated from a centered normalizing constant with explicitly stated Monte Carlo settings.

### G.1. Synthetic design and parameterization

All experiments are conducted in the one-factor Gaussian covariance model

$$\Sigma(\psi, a) = \mathrm{diag}(\psi) + aa^\top, \qquad \psi \in (0,\infty)^p, \quad a \in \mathbb{R}^p.$$

We work in transformed coordinates

$$\eta = \log \psi \in \mathbb{R}^p,$$

so the numerical state variable is $(\eta, a) \in \mathbb{R}^{2p}$ and positivity of the uniqueness parameters is automatic. Unless stated otherwise, the prior on $(\eta, a)$ is standard Gaussian:

$$\eta_i \overset{\text{iid}}{\sim} N(0,1), \qquad a_i \overset{\text{iid}}{\sim} N(0,1).$$

This prior is smooth, everywhere positive, and therefore compatible with the local RLCT analysis.

For the main one-factor experiments in Figs. 1 to 4, we fix $p = 4$ and

$$\psi_0 = (1.0,\ 1.2,\ 0.9,\ 1.1).$$

The three target strata are represented by the loading vectors

$$a_0 = (1.1,\ 0.9,\ 1.0,\ 0.8) \quad \text{(generic)},$$

$$a_0 = (1,\ 1,\ 0,\ 0) \quad \text{(intermediate singular stratum)},$$

and

$$a_0 = (0,\ 0,\ 0,\ 0) \quad \text{(diagonal)}.$$

The target covariance is always $\Sigma_0 = \mathrm{diag}(\psi_0) + a_0 a_0^\top$, and datasets are simulated i.i.d. from $N_p(0, \Sigma_0)$.

For the across-dimension diagnostic in Fig. 5, we vary

$$p \in \{3, 4, 5, 6, 8\},$$

retain the same three one-factor strata, and define $\psi_0(p)$ to be the first $p$ coordinates of the fixed base pattern

$$(1.00,\ 1.20,\ 0.90,\ 1.10,\ 1.05,\ 0.95,\ 1.15,\ 0.85).$$

The generic loading vector is taken to be the equally spaced vector

$$a_0(p) = \big(0.85, \ldots, 1.15\big) \in \mathbb{R}^p,$$

the intermediate singular stratum uses exactly two nonzero coordinates,

$$a_0(p) = (1.0,\ 0.9,\ 0, \ldots, 0),$$

and the diagonal stratum uses $a_0(p) = 0$. These choices match the three asymptotic formulas from Corollary 11.

### G.2. Exact evaluation for the local-comparison plots

Figures 1 and 3 do not require Monte Carlo integration. They are based on exact evaluation of the population excess losses from Lemmas 25 and 26. Writing $(\eta_0, a_0)$ for the target parameter, we generate perturbations

$$(\eta_\rho, a_\rho) = (\eta_0, a_0) + \rho u, \qquad \rho = 2^{-m}, \quad m = 2, \ldots, 8,$$

where $u$ is drawn uniformly from the Euclidean unit sphere in $\mathbb{R}^{2p}$ by normalizing a Gaussian vector. For each perturbation we compute

$$K_{\log}(\theta_\rho), \qquad K_{\mathrm{H}}(\theta_\rho), \qquad \|\Sigma(\theta_\rho) - \Sigma_0\|_{\mathrm{F}}^2 .$$

Figure 1 uses 4000 random directions per stratum and reports medians together with interquartile bands of the ratios

$$\frac{K_{\log}(\theta_\rho)}{\|\Sigma(\theta_\rho) - \Sigma_0\|_{\mathrm{F}}^2}, \qquad \frac{K_{\mathrm{H}}(\theta_\rho)}{\|\Sigma(\theta_\rho) - \Sigma_0\|_{\mathrm{F}}^2}.$$

Figure 3 uses 700 random directions per stratum and plots

$$K_{\mathrm{H}}(\theta_\rho) \quad \text{versus} \quad K_{\log}(\theta_\rho)$$

on log-log axes, with points colored by the locality level $m$. The slope-one guide band is formed from the empirical 10%, 50%, and 90% quantiles of the local ratio $K_{\mathrm{H}}/K_{\log}$ restricted to the smallest radii $\rho \le 2^{-6}$.

### G.3. Centered free-energy estimation by adaptive tempered SMC

Figures 4, 2, and 5 depend on the centered partition function

$$\widetilde{Z}_n^{\ell}(\theta^\star) = \int_\Theta \exp\Big( - \sum_{i=1}^n \big[\ell(X_i, \theta) - \ell(X_i, \theta^\star)\big]\Big)\pi(\theta)\, d\theta,$$

with centered free energy

$$\widetilde{F}_n^{\ell}(\theta^\star) = -\log \widetilde{Z}_n^{\ell}(\theta^\star).$$

In the covariance models considered here, $\theta^\star$ is any representative satisfying $\Sigma(\theta^\star) = \Sigma_0$; the choice of representative is immaterial.

We estimate $\widetilde{Z}_n^\ell$ by adaptive tempered sequential Monte Carlo in the spirit of Del Moral et al. (2006). Let

$$L_{c,n}^\ell(\theta) := \sum_{i=1}^n \big[\ell(X_i,\theta) - \ell(X_i,\theta^\star)\big].$$

The tempering path is

$$\pi_t^\ell(d\theta) \propto \exp\big(-tL_{c,n}^\ell(\theta)\big)\pi(\theta)\,d\theta, \qquad t\in[0,1].$$

Thus $t=0$ corresponds to the prior and $t=1$ corresponds to the centered generalized posterior normalizing constant.

At each temperature update, if the current particle cloud is $\{\theta_i^{(k-1)}\}_{i=1}^N$ and the next temperature is $t_k$, then the incremental weights are

$$w_i^{(k)} = \exp\big(-(t_k - t_{k-1})L_{c,n}^\ell(\theta_i^{(k-1)})\big),$$

and the log-normalizing-constant increment is estimated by

$$\log \widehat{Z}_k^\ell - \log \widehat{Z}_{k-1}^\ell = \log\Big(\frac{1}{N}\sum_{i=1}^N w_i^{(k)}\Big).$$

The temperatures $t_k$ are chosen adaptively so that the effective sample size of the incremental weights stays near a fixed fraction of the particle count, namely $0.82N$ in the reported free-energy experiments. Resampling is systematic. After resampling, particles are rejuvenated by Gaussian random-walk Metropolis steps in the transformed coordinates $(\eta, a)$, with proposal covariance

$$\frac{2.38^2}{d}\,\widehat{\Sigma}_{\mathrm{emp}}, \qquad d = 2p,$$

plus a small ridge term for numerical stability. Because the one-factor covariance map is invariant under the sign flip $a \mapsto -a$, each rejuvenation stage also includes an exact symmetry move that flips the sign of $a$ independently with probability $1/2$.

For the one-factor model, all likelihood-dependent quantities inside the SMC update are evaluated analytically. Writing $D = \mathrm{diag}(e^\eta)$, the covariance matrix is

$$\Sigma(\eta, a) = D + aa^\top,$$

and the Sherman–Morrison formula gives

$$\Sigma(\eta,a)^{-1} = D^{-1} - \frac{D^{-1}aa^\top D^{-1}}{1 + a^\top D^{-1}a}.$$

Likewise, the matrix determinant lemma yields

$$\log\det\Sigma(\eta,a) = \sum_{i=1}^p \eta_i + \log\big(1 + a^\top D^{-1}a\big).$$

These identities are used to compute both centered cumulative losses exactly.

### G.4. Aggregation on the Z-scale and paired design

Repeated normalizing-constant estimates are aggregated on the $Z$-scale rather than the free-energy scale. Concretely, if $\widehat{Z}_{n,1}^\ell, \ldots, \widehat{Z}_{n,M}^\ell$ are repeated SMC estimates for the same dataset, we report

$$\widehat{F}_n^\ell = -\log\Big(\frac{1}{M}\sum_{m=1}^M \widehat{Z}_{n,m}^\ell\Big), \tag{54}$$

not the average of $-\log \widehat{Z}^{\ell}_{n,m}$. Since $-\log(\cdot)$ is convex, (54) is less susceptible to Jensen bias than averaging the free energies directly.

The free-energy experiments also use a paired-data design: for each synthetic dataset, the Gaussian and Hyvärinen centered partition functions are evaluated on exactly the same observations. This is especially important for Fig. 2, where we plot

$$D_n = \widetilde{F}^{\mathrm{H}}_n - \widetilde{F}^{\log}_n.$$

Pairing sharply reduces variance in the difference by removing dataset-to-dataset fluctuations that are common to both losses.

In the across-dimension experiment of Fig. 5, the sample-size grid is dimension-scaled:

$$n = m\,p, \qquad m \in \{400, 800, 1600, 3200\}.$$

Moreover, for each dataset replicate and each fixed $(p, \Sigma_0)$, the values across $n$ are constructed from prefixes of a single long simulated dataset. This nested construction stabilizes slope estimation by ensuring that the free-energy curve across $n$ is monotone in information content rather than being driven by unrelated datasets at each sample size.

### G.5. Slope estimation and uncertainty quantification

All slope estimates are obtained by weighted least squares on the dataset-level mean curves. If $Y_{n,1}, \ldots, Y_{n,R}$ denote the replicated values of a plotted quantity at sample size $n$, we write

$$\bar{Y}_n = \frac{1}{R}\sum_{r=1}^{R} Y_{n,r}, \qquad \widehat{\mathrm{Var}}(\bar{Y}_n) = \frac{s_n^2}{R},$$

where $s_n^2$ is the empirical variance across datasets. Weighted least squares uses weights

$$w_n = \widehat{\mathrm{Var}}(\bar{Y}_n)^{-1}.$$

For Fig. 4, we fit

$$\bar{F}^{\ell}_n = \alpha_\ell + \lambda_\ell \log n + \varepsilon_n$$

over the regime $n \geq 400$. For Fig. 2, we fit

$$\bar{D}_n = \alpha + \beta \log n + \varepsilon_n$$

over the more conservative asymptotic regime $n \geq 1600$. For Fig. 5, we first estimate a slope $\hat{\lambda}(p)$ separately at each dimension $p$, again using weighted least squares, and then plot the residual

$$\hat{\lambda}(p) - \lambda_\star(p)$$

relative to the exact one-factor formula.

All confidence intervals reported in the numerical section are dataset-level bootstrap intervals. In particular, we resample the datasets, not the individual SMC runs or particles. This treats Monte Carlo variability as part of the estimator but preserves the dataset as the basic sampling unit for uncertainty quantification.

### G.6. Experiment settings and reproducibility

Table 1 summarizes the settings used for the reported figures. All random number generation is controlled by a single master seed, from which per-dataset and per-estimator seeds are derived deterministically.

**Table 1.** Numerical settings used for the reported figures. Here $R$ is the number of datasets, $M$ the number of repeated SMC estimates aggregated on the $Z$-scale, and $N$ the particle count for adaptive tempered SMC.

| Fig. | Dim. $p$ | Sample sizes | $R$ | $M$ | $N$ | Notes |
|---|---|---|---|---|---|---|
| 1 | 4 | $\rho = 2^{-m}$, $m = 2, \dots, 8$ | exact | – | – | 4000 rand. dirs. |
| 3 | 4 | $\rho = 2^{-m}$, $m = 2, \dots, 8$ | exact | – | – | 700 rand. dirs. |
| 4 | 4 | $n \in \{100, \dots, 3200\}$ | 24 | 3 | 1800 | fit $n \geq 400$ |
| 2 | 4 | $n \in \{200, \dots, 12800\}$ | 24 | 4 | 2500 | fit $n \geq 1600$ |
| 5 | $3, 4, 5, 6, 8$ | $n = mp$, $m \in \{400, \dots, 3200\}$ | 12 | 2 | 2500 | residual plot |

The appendix figures are intended as diagnostics, not as additional asymptotic theorems. Their role is to demonstrate that the numerical evidence is internally coherent with the proof structure of the paper: exact local quadratic equivalence is visible without Monte Carlo, the raw free-energy curves have the expected $\log n$ scale, and the paired free-energy difference removes the dominant $\lambda \log n$ term exactly as predicted.

## H. Additional Numerical Diagnostics

This appendix collects numerical diagnostics that support, but do not materially sharpen, the main empirical conclusions in Section 6. The figures reported here are intentionally secondary. The main paper uses Fig. 1 to validate the local quadratic comparison from Theorem 7 and Fig. 2 to validate the cancellation of the $\lambda \log n$ term predicted by Theorem 9. By contrast, the present appendix records three additional views of the same numerical story: a direct pointwise comparison of the two excess losses, the raw centered free-energy curves before differencing, and an across-dimension residual diagnostic relative to the exact one-factor formulas. We include these figures for transparency and interpret them conservatively, since they are more sensitive to finite-sample effects and Monte Carlo noise than the two main-paper figures.

### H.1. Direct local comparison of the two excess losses

The first supplementary figure complements Fig. 1. Instead of normalizing each excess loss by $\|\Sigma(\theta) - \Sigma_0\|_{\mathrm{F}}^2$, it plots the pair

$$\big(K_{\log}(\theta),\, K_{\mathrm{H}}(\theta)\big)$$

directly on log-log axes. Since Theorem 3 is driven by local comparability near the common zero fiber, a slope-one relationship between the two losses at sufficiently small perturbation radii is the most literal numerical signature of the theorem. That is exactly what Fig. 3 shows: once the perturbation enters the local regime, the cloud tightens around a narrow slope-one band in all three one-factor strata.

The figure should not be over-interpreted as a separate theorem. It is simply a geometric restatement of the same phenomenon already visible in Fig. 1. The value of reporting it is diagnostic: it makes the bounded-ratio relation $K_{\mathrm{H}} \asymp_{\mathcal{M}(\Sigma_0)} K_{\log}$ visible directly, without reference to the covariance norm.

### H.2. Raw centered free-energy curves

The second supplementary figure shows the centered free energies

$$\widetilde{F}_n^{\log} \qquad \text{and} \qquad \widetilde{F}_n^{\mathrm{H}}$$

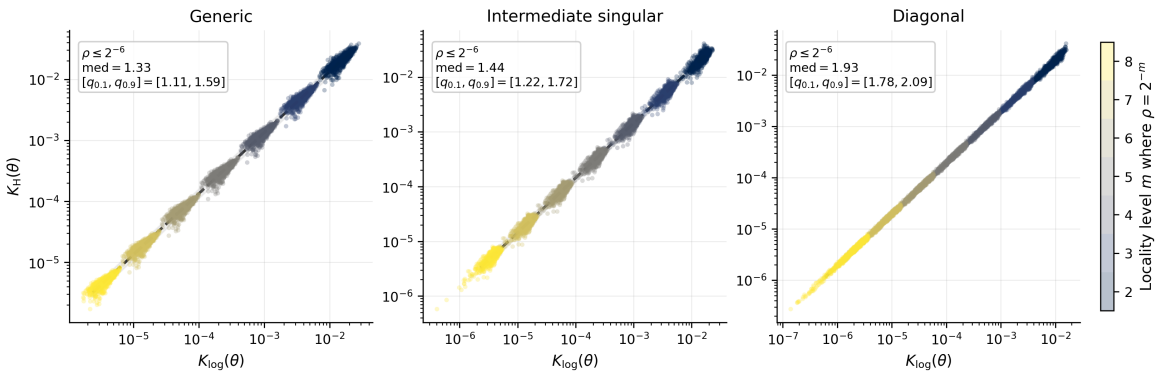


**Figure 3. Direct local comparison between the Gaussian and Hyvärinen excess losses.** For each one-factor stratum, we perturb the target parameter in transformed coordinates $(\eta, a)$ and plot $K_{\mathrm{H}}(\theta)$ against $K_{\log}(\theta)$ on log-log axes. Points are colored by locality level $m$, where $\rho = 2^{-m}$. The dashed line is the slope-one guide $K_{\mathrm{H}} = c\,K_{\log}$, where $c$ is the empirical median ratio over the smallest radii $\rho \le 2^{-6}$; the shaded band corresponds to the local 10%–90% quantile range of $K_{\mathrm{H}}/K_{\log}$. The narrow slope-one structure in the local regime is consistent with the local comparison relation underlying Theorem 3.

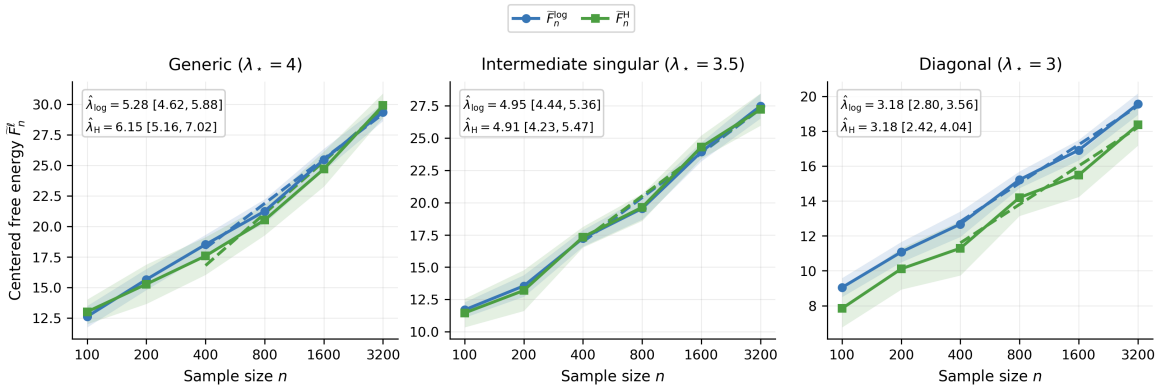


**Figure 4. Raw centered free-energy curves before differencing.** Dataset means of the centered free energies $\widetilde{F}_n^{\log}$ and $\widetilde{F}_n^{\mathrm{H}}$ are plotted against sample size $n$, with pointwise 95% confidence bands and large-$n$ fitted lines. The horizontal axis is logarithmic in $n$. The curves display the expected $\log n$-type growth and stratum ordering, but finite-sample and Monte Carlo effects remain visible. Accordingly, this figure is interpreted as a qualitative diagnostic, while the paired-difference plot in Fig. 2 serves as the main empirical transfer check.

before differencing. In principle, one could attempt to validate the transfer theorem directly from these raw curves by comparing fitted slopes with the theoretical one-factor values from Corollary 11. In practice, however, this is a substantially noisier diagnostic than the paired-difference plot in the main text. The reason is structural: each curve contains both the leading $\lambda \log n$ term and a non-negligible $O_p(1)$ offset, and finite-$n$ Monte Carlo fluctuations affect the two losses in correlated but not identical ways.

For this reason, Fig. 4 is included only as a qualitative support figure. It confirms that the centered free energies are approximately linear in $\log n$ and that the three strata exhibit the expected ordering of asymptotic difficulty, but it does not provide the sharpest transfer check. The paired difference in Fig. 2 is the more informative object because it removes the dominant $\lambda \log n$ term by design.

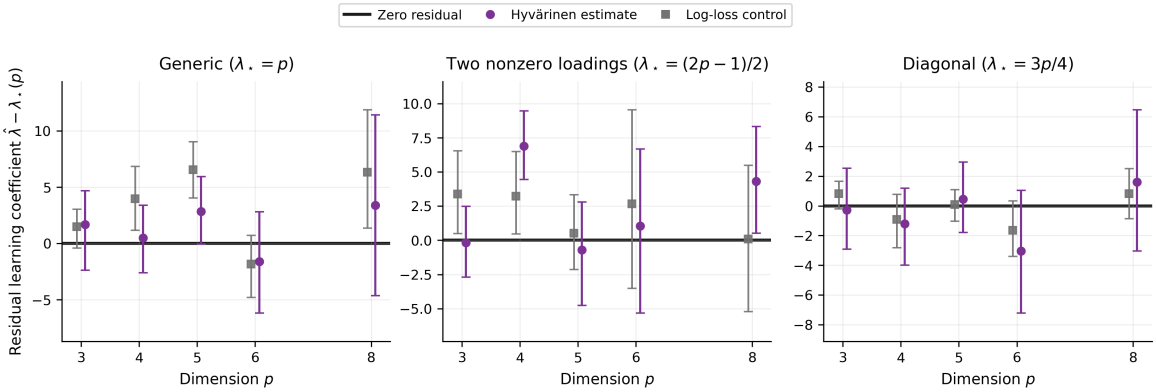


**Figure 5. Across-dimension residual slopes relative to the exact one-factor formulas.** For each stratum and each ambient dimension $p$, the plotted quantity is the residual learning-coefficient estimate $\hat{\lambda} - \lambda_\star(p)$, where $\lambda_\star(p)$ is the exact one-factor formula from Corollary 11. Purple circles correspond to Hyvärinen estimates, gray squares to Gaussian log-loss controls, and the horizontal zero line is the exact transferred target. The figure is diagnostic rather than load-bearing: uncertainty remains substantial at larger $p$, but the Hyvärinen and Gaussian estimates display broadly similar deviation patterns across dimensions.

### H.3. Across-dimension residual slopes

The final supplementary figure studies the one-factor exact formulas across ambient dimension $p$. The most direct raw-slope version of this experiment proved too sensitive to finite-$n$ and finite-SMC bias, so the reported figure centers each estimated slope by the exact transferred formula

$$\lambda_\star(p) \in \left\{ p,\ \frac{2p-1}{2},\ \frac{3p}{4} \right\},$$

depending on the stratum. The plotted quantity is therefore

$$\hat{\lambda}(p) - \lambda_\star(p),$$

with zero corresponding to exact agreement with the asymptotic theory.

We emphasize that Fig. 5 is diagnostic rather than load-bearing. Its role is not to provide a new asymptotic theorem, but to show that the Hyvärinen and Gaussian slope estimates exhibit broadly similar finite-sample deviation patterns as $p$ varies. This is the correct standard for an auxiliary across-dimension check in the present paper, because the new contribution is a transfer theorem between losses, not an independent finite-sample convergence theorem for slope estimators. The exact asymptotic formulas themselves are imported from the singular-learning-theory analysis of Drton et al. (2025) and transferred here via Theorem 9.

**Summary.** Taken together, the supplementary diagnostics reinforce the interpretation of the main text. Figure 3 confirms the pointwise local-comparison structure, Fig. 4 shows the raw $\log n$-scale growth before differencing, and Fig. 5 indicates that the transferred one-factor formulas remain qualitatively visible across dimensions. None of these figures is individually as decisive as Figs. 1 and 2, but together they provide a fuller picture of the numerical behavior of the Gaussian and Hyvärinen generalized posteriors in the singular one-factor model.