SHARP VARIANCE CERTIFICATES FOR LOG-PARTITION GAPS

Anonymous authors
Paper under double-blind review

ABSTRACT

When does the variance of a log density ratio certify, rather than merely approximate, the error of a Gaussian approximation? For a standard Gaussian reference Q and a smooth log ratio g that is square-integrable with its gradient, we prove that $\nabla^2 g \preceq (1-\kappa)I$, $0 < \kappa \leq 1$, implies $\mathrm{KL}(Q\|P) = \log \mathbb{E}_Q e^g - \mathbb{E}_Q g \leq \mathrm{Var}_Q(g)/(2\kappa)$ for the tilted target $P \propto e^g Q$. The coefficient is optimal for every κ, and the bound extends the known concave case to nonconcave log ratios without bounded-gradient or small-error assumptions. A matrix refinement resolves direction and Hermite degree. It never exceeds the population matrix Fisher (log-Sobolev) certificate, is exact for Gaussian targets when the curvature matrix is their relative precision, and shows that weak curvature is paid only through the degree-one and degree-two components. These are the stationarity residuals of Gaussian variational inference and are known to vanish at its exact optimum. Consequently, at an exact full-covariance VI stationary point of a smooth log-concave target, $\mathrm{KL}(Q\|P) \leq 3\,\mathrm{Var}_Q(g)/2$ under mild growth conditions, with no curvature margin. Explicit examples separate the variance, degree-resolved and Fisher certificates at nonvanishing error, and the bound extends to non-Gaussian references with two-sided curvature control. Gaussian polynomial moments and residual bounds make the certificates computable without the target normalizer. On Bayesian logistic posteriors, the scalar certificate is 0.29–0.67 times matrix Fisher across 48 historical fits, and under strong regularization it certifies KL tolerances that the Fisher certificate cannot reach. Under a fixed prior, the degree-resolved certificate decreases as data accumulate, and independent estimates place it within a factor of two of the estimated KL at moderate priors.

1 INTRODUCTION

Laplace approximations and variational inference (VI) replace an intractable posterior P by a Gaussian Q, but do not say how far Q is from P. The error $\mathrm{KL}(Q\|P)$ is the gap between $\log Z_P$ and the evidence lower bound (ELBO), so computing it requires the normalizer that approximate inference avoids. When can that error be *certified*, that is, bounded above using quantities computed under Q alone?

A natural candidate is the variance under Q of the log density ratio $g = \log(\widetilde{p}/q)$: the unknown normalizer only shifts g, and half the variance is the leading term of KL for accurate approximations (Dehaene, 2019). But a leading term is not a bound, and without further structure no multiple of the variance bounds KL (Appendix E). The same quantity is also a training objective: log-variance losses for variational inference and neural samplers minimize it (Richter et al., 2020; Nüsken & Richter, 2021; Richter & Berner, 2024). When the variance is taken under Q itself, a certificate $\mathrm{KL}(Q\|P) \leq C\ \mathrm{Var}_Q(g)$ says when a small population value of such a loss guarantees a small error for a Gaussian Q.

Our answer is *relative curvature*: in general, the target must be at least a fraction κ as sharply curved as the Gaussian, in every direction and at every point. If $Q = N(\mu, \Sigma)$ and $\nabla^2(-\log \widetilde{p}) \succeq \kappa\Sigma^{-1}$ everywhere, with $0 < \kappa \leq 1$, then

$$\mathrm{KL}(Q\|P) \leq \frac{1}{2\kappa}\,\mathrm{Var}_Q\left(\log\frac{\widetilde{p}}{q}\right),$$

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

and the constant $1/(2\kappa)$ cannot be improved for any κ. With log-concave likelihoods, a Gaussian prior supplies this curvature.

Contributions.

- **A sharp variance certificate** (Theorem 3.1). The bound allows log density ratios that are not concave and have unbounded gradients, and needs no small-error expansion. The concave case $\kappa = 1$ is known (Chen, 2026). We cover $0 < \kappa < 1$ with the optimal constant and extend the bound, with a sharp factor $\beta/(2(\alpha - \rho))$, to non-Gaussian references with two-sided curvature control (Theorem C.1).
- **A degree-resolved certificate and exact Gaussian VI** (Theorem 3.2, Corollaries 3.3 and 3.4). The matrix refinement retains both direction and Hermite degree, never exceeds matrix Fisher at the population level, and equals the true KL for Gaussian targets with matching relative precision. Weak curvature is paid only through the degree-one and degree-two components, which are the whitened gradients of the Gaussian VI objective; each higher degree has coefficient at most $3/2$. These components vanish at an exact full-covariance VI stationary point (Katsevich & Rigollet, 2024), and we show that $\mathrm{KL}(Q\|P) \le 3\,\mathrm{Var}_Q(g)/2$ then holds for smooth log-concave targets with no curvature margin.
- **Separating examples.** On a smooth target the variance certificate stays bounded while the Fisher certificate diverges. On an exactly fitted VI example, as $\kappa \downarrow 0$, the scalar certificate diverges while the degree-resolved one tends to about 3.5 times the true KL.
- **A computable implementation and experiments.** Gaussian polynomial moments and explicit residual bounds evaluate the certificates without the target normalizer. On 84 planned logistic-regression approximations (48 historical and 36 prospective, with the three non-returned fits counted as failures), the scalar variance certificate is 0.29–0.67 times matrix Fisher on the historical fits (median 0.43), the degree-resolved certificate is smaller than the variance certificate on every returned fit, and both certify tolerances that the Fisher certificate cannot reach under strong regularization. Independent estimates place the certificate within 1.3–1.8 times the estimated KL at moderate priors, and under a genuinely fixed prior it keeps decreasing as data accumulate, although the global curvature margin shrinks.

The ELBO identity and the log-variance diagnostic are established; our contribution is the guarantee, its refinements and their computation.

2 Setup: the ELBO gap and what a certificate needs

Let Q be the approximating distribution with density q, and let $\widetilde{p}$ be an unnormalized target density with normalizer $Z_P = \int \widetilde{p}$ and normalized law P. Write $g = \log(\widetilde{p}/q)$ for the log density ratio (log ratio for short). More generally, for any $g \in L^2(Q)$ with $0 < \mathbb{E}_Q e^g < \infty$, define

$$J_Q(g) = \log \mathbb{E}_Q e^g - \mathbb{E}_Q g, \qquad V_Q(g) = \mathrm{Var}_Q(g), \qquad \frac{dP_g}{dQ} = \frac{e^g}{\mathbb{E}_Q e^g}. \tag{1}$$

The standard variational identity (Blei et al., 2017) reads

$$\log Z_P = \log \mathbb{E}_Q e^g = \underbrace{\mathbb{E}_Q g}_{\text{ELBO}} + J_Q(g), \qquad J_Q(g) = \mathrm{KL}(Q\|P_g). \tag{2}$$

Adding a constant to g changes neither J_Q nor V_Q, so V_Q can be computed without Z_P.

A certificate needs the geometric inequality $J_Q(g) \le C\, V_Q(g)$ (Sections 3–4) and a guaranteed upper bound on the population variance (Section 5).

Whitening and relative curvature. For $Q = N(\mu, \Sigma)$, the coordinates $z = \Sigma^{-1/2}(x - \mu)$ make Q standard normal. The condition $\nabla^2(-\log \widetilde{p}) \succeq \kappa\Sigma^{-1}$ then becomes $\nabla^2 g \preceq (1 - \kappa)I$: everywhere and in every direction, the target is at least κ times as curved as the approximation's precision (a more curved target can use $\kappa = 1$). The condition is global; a Hessian at the mode does not verify it. For Bayesian logistic regression with prior precision Λ and $Q = N(\mu, LL^\top)$, $\kappa = \min\{1, \lambda_{\min}(L^\top \Lambda L)\}$ is valid: the prior's share of the approximation's precision in its weakest direction. With a fixed prior it shrinks like $1/n$.

3 VARIANCE CERTIFICATES UNDER RELATIVE CURVATURE

3.1 THE SCALAR CERTIFICATE

Theorem 3.1 (Gaussian curvature–variance certificate). *Let $Q = N(\mu, \Sigma)$, $\Sigma \succ 0$, with density q, and let $\widetilde{p} : \mathbb{R}^d \to (0, \infty)$ be C^2. Set $g = \log(\widetilde{p}/q)$. Suppose*

$$g,\ \Sigma^{1/2}\nabla g \in L^2(Q), \qquad \nabla^2(-\log \widetilde{p})(x) \succeq \kappa\Sigma^{-1} \quad \textit{for every } x, \qquad 0 < \kappa \le 1.$$

Then $\widetilde{p}$ has a finite positive integral and its normalized law P satisfies

$$\mathrm{KL}(Q\|P) = J_Q(g) \le \frac{V_Q(g)}{2\kappa}. \tag{3}$$

The coefficient $1/(2\kappa)$ is optimal for each fixed $\kappa \in (0, 1]$, already in dimension one.

In whitened coordinates the assumptions read $g \in C^2 \cap W^{1,2}(Q)$ (g and ∇g square integrable) and $\nabla^2 g \preceq (1-\kappa)I$. For $\kappa < 1$ the log ratio may be nonconcave; no bounded gradients, density-ratio bounds or small errors are assumed. At $\kappa = 1$ this is a special case of Chen (2026, Theorem 2(1)), which needs less regularity. Once $V_Q(g) \le v$ is known, the certificate also brackets the normalizer: $\mathbb{E}_Q g \le \log Z_P \le \mathbb{E}_Q g + v/(2\kappa)$.

Why variance appears. The proof uses a standard stochastic-control representation of the log normalizer (Boué & Dupuis, 1998; Borell, 2000), in the version of Hariya & Watanabe (2022, Proposition 3.1) for unbounded functions. It trades steering a Brownian motion W on $[0, 1]$ toward large g against steering energy. Let $\mathcal{A}$ be the adapted drifts with finite expected energy, $h_u = \int_0^1 u_t\, dt$, and $m_t = \mathbb{E}[\nabla g(W_1) \mid \mathcal{F}_t]$. Then

$$\begin{aligned} J_Q(g) &= \sup_{u\in\mathcal{A}} \mathbb{E}\left[g(W_1 + h_u) - g(W_1) - \frac{1}{2}\int_0^1 \|u_t\|^2\, dt\right] \\ &\le \sup_{u\in\mathcal{A}} \mathbb{E}\int_0^1 \left[m_t^\top u_t - \frac{\kappa}{2}\|u_t\|^2\right] dt \le \frac{1}{2\kappa}\mathbb{E}\int_0^1 \|m_t\|^2\, dt = \frac{V_Q(g)}{2\kappa}. \end{aligned}$$

The curvature bound caps the gain from steering; completing the square and the Clark–Ocone identity turn the conditional gradient energy into the variance. Appendix A gives the integrability details. At $\kappa = 1$ the same argument recovers Chen's endpoint in this smooth class.

Sharpness and boundaries. For $Q = N(0, 1)$, $g_b(z) = (1-\kappa)z^2/2 + bz$ has $J_Q(g_b) = [-\log\kappa - (1-\kappa)]/2 + b^2/(2\kappa)$ and $V_Q(g_b) = (1-\kappa)^2/2 + b^2$, so the ratio tends to $1/(2\kappa)$ as $|b| \to \infty$. Affine log ratios attain equality at $\kappa = 1$. No finite constant works at $\kappa = 0$ without further structure (compare Corollary 3.4), and the same constant fails for $\kappa > 1$ (Appendix E). Appendix A.1 gives an equivalent one-sided Bernstein form.

Comparison with Fisher certificates. The classical Bakry–Émery logarithmic Sobolev inequality gives $J_Q(g) \le I_Q(g)/(2\kappa)$, where $I_Q(g) = \mathbb{E}_Q\|\nabla g\|^2$ in whitened coordinates (Bakry & Émery, 1985). Write $g = \sum_{k\ge 0} g_k$ in Hermite components of total degree k, and let $V_k = \mathrm{Var}_Q(g_k)$. Then $V_Q(g) = \sum_{k\ge 1} V_k$ and $I_Q(g) = \sum_{k\ge 1} kV_k$. Thus, for the same scalar curvature margin, the variance certificate is no larger than scalar Fisher; when the variance is nonzero, their ratio is determined by the variance-weighted mean Hermite degree. A stronger comparator retains a positive definite curvature matrix A satisfying $\nabla^2 g \preceq I - A$ and gives $F_A(g) = \frac{1}{2}\mathbb{E}_Q[\nabla g^\top A^{-1}\nabla g]$. The scalar variance certificate is not generally ordered with F_A: replacing A by one worst-direction margin discards directional information. Theorem 3.2 retains both degree and direction and is no larger than matrix Fisher at the population level.

3.2 RESOLVING DEGREE AND DIRECTION

Theorem 3.2 (Degree-resolved certificate). *Let $Q = N(0, I)$, $g \in C^2 \cap W^{1,2}(Q)$ with Hermite components g_k, and suppose $\nabla^2 g \preceq I - K$ everywhere for a constant symmetric $K \succ 0$. Then*

$$J_Q(g) \le \frac{1}{2}\sum_{k\ge 1} \mathbb{E}_Q[\nabla g_k^\top W_k(K)\nabla g_k], \qquad W_k(K) = \int_0^1 t^{k-1}[K + (I-K)t]^{-2}\, dt. \tag{4}$$

Equality holds for every Gaussian target, $g(z) = z^\top(I-K)z/2 + b^\top z + c$. The right side is at most both the matrix Fisher certificate $F_K(g)$ and $V_Q(g)/(2\min\{1, \lambda_{\min}(K)\})$.

Theorem 3.2 is proved as Theorem B.1. It holds for every positive definite K, with no upper restriction, and its weights show where weak curvature costs anything. Equality for a Gaussian target uses its exact relative precision as K; this equality does not assert separate optimality of every higher-degree weight.

Corollary 3.3 (Where weak curvature is paid)**.** *Under the assumptions of Theorem 3.1, in whitened form, with $V_k = \mathrm{Var}_Q(g_k)$,*

$$J_Q(g) \le \frac{V_1}{2\kappa} + W_2(\kappa)V_2 + \frac{3}{2}\sum_{k\ge 3} V_k, \qquad W_2(\kappa) = \frac{\kappa - 1 - \log\kappa}{(1-\kappa)^2}, \quad W_2(1) = \tfrac{1}{2}. \tag{5}$$

As $\kappa \to 0$, $W_2(\kappa)$ grows like $\log(1/\kappa)$. If g is even and $g_2 = 0$, then $J_Q(g) \le V_Q(g)$ (proof in Appendix B.1).

Consequence for Gaussian VI. Up to sign and whitening, the two low-degree components are the gradients of the VI objective. For $Q = N(\mu, LL^\top)$, let $\mathcal{F}(\mu, L) = -\mathbb{E}_Q \log \widetilde{p} - H(Q)$ be the reverse-KL objective up to a constant. Under the growth conditions of Lemma B.2, Gaussian integration by parts gives $g_1(z) = a^\top z$ and $g_2(z) = \frac{1}{2}(z^\top B z - \operatorname{tr} B)$, with $a = \mathbb{E}_Q \nabla g = -L^\top \nabla_\mu \mathcal{F}$ and $B = \mathbb{E}_Q \nabla^2 g = -L^\top \nabla_L \mathcal{F}$. Thus $V_1 = \|a\|^2$, $V_2 = \frac{1}{2}\|B\|_F^2$, and Corollary 3.3 reads

$$J_Q(g) \le \frac{\|a\|^2}{2\kappa} + \frac{W_2(\kappa)}{2}\|B\|_F^2 + \frac{3}{2}\sum_{k\ge 3} V_k. \tag{6}$$

Weak curvature is paid only through the stationarity residual of the fitted Gaussian; the non-Gaussian part of the error has a curvature-free coefficient. For logistic regression with a fixed prior, the certified κ shrinks like $1/n$ (Section 2), but (6) charges this only against a and B, which an accurate VI solver drives to zero. At an exact stationary point, $a = B = 0$ (the cancellation used by Katsevich & Rigollet, 2024, Section 4.2). If the target is also log-concave, a limiting argument removes the curvature margin altogether.

Corollary 3.4 (Exact Gaussian VI on log-concave targets)**.** *Let $\widetilde{p}$ be log-concave, with the growth conditions of Lemma B.2, and let Q be an exact interior stationary point of the reverse-KL objective over the full Gaussian family. Then $\widetilde{p}$ is normalizable and*

$$\mathrm{KL}(Q\|P) \le \frac{3}{2}\mathrm{Var}_Q(g), \tag{7}$$

with coefficient 1 *when the whitened log ratio is even. No positive curvature margin is required.*

The proof applies Corollary 3.3 to θg, $\theta < 1$, and lets $\theta \uparrow 1$ (Appendix B.1). Proposition E.1 shows that the cancellation is essential: without it, no constant works at $\kappa = 0$. A solver termination flag, or stationarity in a restricted Gaussian family, does not establish the cancellation; for numerical fits we therefore evaluate the measured low-degree terms.

4 What each certificate can see: separating examples

The following examples separate gradient energy (Fisher), variance (Theorem 3.1) and its degree split (Theorem 3.2) at fixed, nonvanishing error.

Variance versus gradient energy: a remote wall. For $Q = N(0,1)$, add a steep but rarely visited convex wall to a quadratic log ratio:

$$g_t(z) = z + \tfrac{1}{4}z^2 - a_t(z-t)_+, \qquad a_t = \sqrt{t/\Pr(Z > t)}, \quad Z \sim Q.$$

The target has curvature $\kappa = 1/2$; the rarely visited wall creates very large gradients. Exact calculations give

$$J_Q(g_t) \to \frac{3}{4} + \frac{\log 2}{2}, \qquad \frac{V_Q(g_t)}{2\kappa} \to \frac{9}{8}, \qquad \frac{I_Q(g_t)}{2\kappa} = t + \frac{5}{4} + o(1). \tag{8}$$

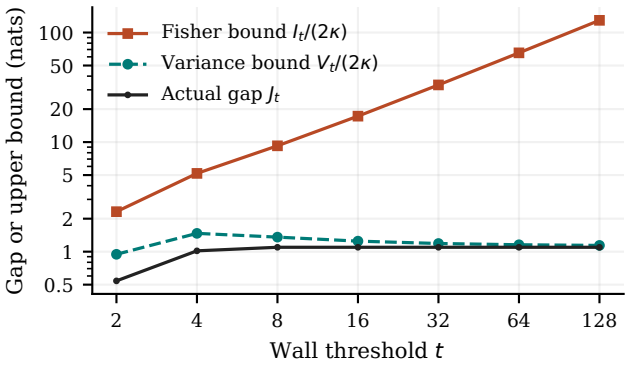


Figure 1: A remote convex wall. The target and the curvature constant are the same for both certificates. The variance certificate stays bounded (limit 1.125) and close to the true KL (limit 1.0966), while the Fisher certificate diverges. Both axes are logarithmic.

The true KL tends to 1.097 and the variance certificate to 1.125, while the Fisher certificate grows without bound (Figure 1). A smooth softplus version of the wall has the same limits, and its Fisher certificate is asymptotic to t (Propositions D.1 and D.2). Its rare-tail fourth moments also make empirical variances unreliable as upper bounds, which is why we bound variances analytically. The example separates variance from Fisher certification; a direct Gaussian envelope for the normalizer is also asymptotically exact on this family.

Degree-resolved versus scalar: an exactly fitted quartic. For $Q = N(0,1)$ and $0 < \kappa < 1$, let

$$g_\kappa(z) = -\frac{1-\kappa}{12} H_4(z), \quad H_4(z) = z^4 - 6z^2 + 3, \qquad P_\kappa(dz) \propto \exp\left[-\frac{\kappa}{2}z^2 - \frac{1-\kappa}{12}z^4\right] dz.$$

Here $g_\kappa'' = (1-\kappa)(1-z^2) \le 1-\kappa$, so the curvature condition holds with margin κ. Moreover, Q is the *exact* Gaussian VI approximation of P_κ: the reverse-KL objective is strictly convex, and its unique stationary point has mean 0 and variance 1. The log ratio is pure degree four, with $V_Q(g_\kappa) = (1-\kappa)^2/6$. Hence the scalar certificate $(1-\kappa)^2/(12\kappa)$ diverges as $\kappa \downarrow 0$, whereas the degree-resolved certificate $(1-\kappa)^2 W_4(\kappa)/3 \le (1-\kappa)^2/6$ tends to $1/6$, and the true KL tends to $J_0 = \log[12^{1/4}\Gamma(1/4)/(2\sqrt{2\pi})] - \frac{1}{4} \approx 0.047$. As $\kappa \downarrow 0$, the ratio of the degree-resolved certificate to the true KL tends to about 3.5, while the true error tends to a positive constant. At the endpoint $\kappa = 0$ the target is still log-concave and Q is still its exact VI fit, so Corollary 3.4 applies directly and gives the same value $1/6$ (Proposition B.3).

5 Computing a certificate

A certificate needs a global curvature matrix and a guaranteed upper bound on the variance or degree-resolved energy of g. Figure 2 replaces g by a polynomial with exact Gaussian moments and pays for the replacement with a residual bound.

Why the residual bound is valid. Centering and the triangle inequality give $\mathrm{Var}_Q(g) \le (\sqrt{\mathrm{Var}_Q(h)} + r_v)^2$ for $g = h + r$; since $\mathcal{E}_K^{1/2}$ is a seminorm, the same argument gives B_D (Appendix B.2). Neither step assumes that the coefficients of h are exact projections. When the stationarity residuals $a = \mathbb{E}_Q \nabla g$ and $B = \mathbb{E}_Q \nabla^2 g$ of g itself are enclosed, as sums of one-dimensional integrals for generalized linear models, the weak-curvature weights act only on them and the polynomial residual is charged with a curvature-free constant (Proposition B.4; used in the diagnostics of Section 6.4).

Structured models. In Bayesian logistic regression, the negative log posterior is a quadratic prior plus one softplus term per observation, each a function of a single linear projection $x_i^\top \theta$. We replace each softplus term by a univariate Taylor or Hermite polynomial and bound its squared residual by a global derivative remainder or on a finite interval with analytic Gaussian tails; the one-dimensional bounds combine by the triangle inequality even though the projections are dependent. Variances of sums of such polynomials need pairwise overlaps $u_i^\top u_j$ of the projection directions, so the cost grows with the number of pairs of likelihood terms even though the theorems' constants are dimension-free.

Certifying a fixed Gaussian approximation. Input: an unnormalized target $\widetilde{p}$, a fixed $Q = N(\mu, LL^\top)$, and a tolerance ε.

1. **Whiten and bound curvature.** Write $g(z) = \log \widetilde{p}(\mu + Lz) + \|z\|^2/2$, up to an additive constant. Find a constant $K \succ 0$ with $\nabla^2 g \preceq I - K$ everywhere. For a Gaussian prior of precision Λ and a log-concave likelihood, use $K = L^\top \Lambda L$. Set $\kappa = \min\{1, \lambda_{\min}(K)\}$. All expectations below are under the whitened standard Gaussian.
2. **Choose a polynomial.** Approximate g by a polynomial h. Its coefficients may be chosen by quadrature, Taylor expansion, or another heuristic.
3. **Evaluate polynomial moments.** Compute $\mathrm{Var}_Q(h)$ and the degree-resolved energy $\mathcal{E}_K(h) = \sum_{k\geq 1} \mathbb{E}_Q[\nabla h_k^\top W_k(K) \nabla h_k]$. Retain the degree-one and degree-two terms as measured.
4. **Bound the actual residual.** For $r = g - h$, obtain $\|r\|_{L^2(Q)} \leq r_v$ and, when used, $\|K^{-1/2}\nabla r\|_{L^2(Q)} \leq r_f$. These bounds must refer to the same polynomial whose moments were evaluated.
5. **Return the bound and its arithmetic-assurance status.** Use the smaller of
$$B_V = (\sqrt{\mathrm{Var}_Q(h)} + r_v)^2/(2\kappa) \quad \text{and} \quad B_D = \tfrac{1}{2}[\sqrt{\mathcal{E}_K(h)} + \min\{r_v/\sqrt{\kappa}, r_f\}]^2$$
(Appendix B.2). When the gradient residual bound is unavailable, use $r_v/\sqrt{\kappa}$ for the residual term. Compare the result with ε and, where needed, convert it to a posterior-mass bound using Appendix F.2. Numerical tolerance decisions retain the assurance status of their evaluation.

Figure 2: The certification procedure used in Section 6.

On six $g_{\mathrm{p}} = 1$ VI fits selected for the stochastic-certifier comparison, the deterministic computation took about 1–4 CPU seconds per fit, including the paired Fisher computation (Appendix F.7); this timing is not representative of every case in the complete grids.

The Fisher comparison and numerical assurance. Matrix Fisher receives the same polynomial choices and residual treatment. Reverse triangle inequalities also give lower endpoints for the Fisher functional: if such an endpoint exceeds a tolerance, no more accurate evaluation of that functional can certify it (this is not a lower bound on the true KL). The logistic results are ordinary floating-point evaluations of the analytic bounds. Wider-arithmetic recomputation of the passing decisions with the smallest margins (Appendices F.4 and F.6) changes no outcome; these are stability checks, not verification of every decision. The Poisson calculations use directed arithmetic (Appendix F.7).

6 Certifying fitted logistic-regression posteriors

Given an already fitted Gaussian, can we certify that $\mathrm{KL}(Q\|P)$ is below a tolerance without the posterior's normalizer? The variance, degree-resolved and matrix-Fisher certificates use the same fitted Gaussians, curvature matrix, polynomials and residual bounds. Appendix F gives the complete procedure. All logistic numbers are ordinary floating-point evaluations of the analytic bounds (Section 5).

6.1 Models and protocol

We use Bayesian logistic regression on Spambase ($d = 58$) and on Covertype class 2 versus the rest ($d = 46$–52, the identified predictor span) (Hopkins et al., 1999; Blackard, 1998). Each dataset supplies sample sizes $n \in \{1024, 4096\}$, three random subsets, and two prior strengths. The design is standardized so that $X^\top X = nI$, and the Gaussian prior has precision $\Lambda = (n/g_{\mathrm{p}})I$, a Zellner-type prior with scale g_{p}. For $g_{\mathrm{p}} = 1$ and 4 the prior supplies at least 80% and 50% of the posterior curvature, a strongly regularized regime. Laplace and full-covariance Gaussian VI fits give 48 fixed approximations. These *historical* grids were exploratory: earlier Banknote and WDBC studies showed no robust threshold advantage, after which we narrowed the prior grid (Appendix F.7).

A *prospective* study then fixed 36 further approximations, before fitting, on three disjoint Covertype blocks. It repeats the settings and adds weaker priors at $n = 4096$: $g_{\mathrm{p}} = 16$ (prior share of curvature at least 20%) and unit prior variance ($g_{\mathrm{p}} = n$). The three unit-prior VI fits reached the fitting cap and count as failures to certify. Cases share data and are not independent.

Table 1: **Certificate ratios for all 81 returned approximations: 48 historical and 33 prospective.** Ratios below one favor the numerator. Matrix Fisher uses $A = L^\top \Lambda L$ and the same polynomial and residual treatment. "Degree" is the degree-resolved certificate in the scalar-summary form (26), evaluated in a reanalysis developed after the original outcomes. The last column reports scalar variance bounds in nats; fit counts are historical + prospective. The three unit-prior VI fits did not return and remain in the denominators of Table 2.

Prior g_{p}	Fit	Fits	Variance/Fisher	Degree/Fisher	Variance cert. (nats)
1	Laplace	12+6	0.48–0.56	0.45–0.53	0.0009–0.0044
1	VI	12+6	0.29–0.33	0.25–0.29	0.0003–0.0012
4	Laplace	12+6	0.58–0.67	0.47–0.52	0.069–0.24
4	VI	12+6	0.32–0.37	0.22–0.25	0.024–0.073
16	Laplace	0+3	0.90–0.91	0.47–0.49	1.97–2.44
16	VI	0+3	0.49–0.51	0.18	0.64–0.81
n (unit variance)	Laplace	0+3	35–40	0.61–0.64	$(1.9\text{–}2.5) \times 10^6$

Table 2: **Certification decisions on the complete logistic grids.** "Excluded" counts variance passes for which a lower endpoint on the Fisher functional excludes the stated tolerance or posterior-mass conversion; it is not a lower bound on KL and need not equal the difference of the pass counts. All 48 historical and 36 planned prospective cases remain in the denominators, including the three non-returned prospective fits. $E_{0.995}$ is the fitted Gaussian's 99.5% ellipsoid.

	Historical /48			Prospective /36		
Decision	Variance	Fisher	Excluded	Variance	Fisher	Excluded
$\mathrm{KL} \le 0.001$	11	1	10	6	0	6
$\mathrm{KL} \le 0.01$	24	24	0	12	12	0
$\mathrm{KL} \le 0.05$	33	24	9	17	12	5
$\mathrm{KL} \le 0.1$	41	30	7	21	15	3
$P(E_{0.995}) \ge 0.95$	30	24	6	15	12	3

6.2 How much tighter are the certificates?

Table 1 compares the certificates group by group. On the 48 historical fits, the scalar variance certificate is 0.29–0.67 times the matrix-Fisher certificate, with median 0.43. It is below even the lower endpoint of the Fisher functional in every historical case, so the gap is not explained by a looser evaluation of that functional. The pattern is consistent with the Hermite structure of Section 3: VI fits have small measured degree-one and degree-two components and ratios near one third, while Laplace ratios are 0.49–0.67; the prospective fits show the same pattern.

As the prior weakens, the worst-direction scalar margin becomes costly, and the degree-resolved certificate recovers the loss. At $g_{\mathrm{p}} = 16$, the Laplace variance/Fisher ratio is 0.90–0.91, while the degree-resolved ratio is 0.47–0.49. With unit prior variance, the scalar certificate is 35–40 times matrix Fisher and the degree-resolved one is 0.61–0.64 times, although all absolute bounds are then vacuous. For the returned VI fits at $g_{\mathrm{p}} = 1, 4, 16$, the degree-resolved certificate is 0.18–0.29 times matrix Fisher, consistent with the mechanism of Corollary 3.3.

6.3 Certification decisions

A tighter bound changes a decision only when the tolerance lies between the evaluated certificates. Table 2 reports decisions at fixed tolerances. At KL tolerance 0.001, variance passes on 11 of 48 historical approximations and 6 of 36 planned prospective cases, compared with 1 and 0 for matrix Fisher; for every one of these additional cases, the lower endpoint of the Fisher functional already exceeds the tolerance. All these passes use $g_{\mathrm{p}} = 1$. At tolerance 0.01, the methods tie: every $g_{\mathrm{p}} = 1$ fit passes under both certificates and every $g_{\mathrm{p}} = 4$ fit fails under both.

A KL bound also controls probabilities of events. If $\mathrm{KL}(Q\|P) \le B$, binary-KL data processing converts a known $Q(E)$ into a lower bound on $P(E)$. Figure 3 applies this to the fitted Gaussian's 99.5% ellipsoid. For all six historical $n = 4096$, $g_{\mathrm{p}} = 4$ VI fits, the variance bounds give posterior-

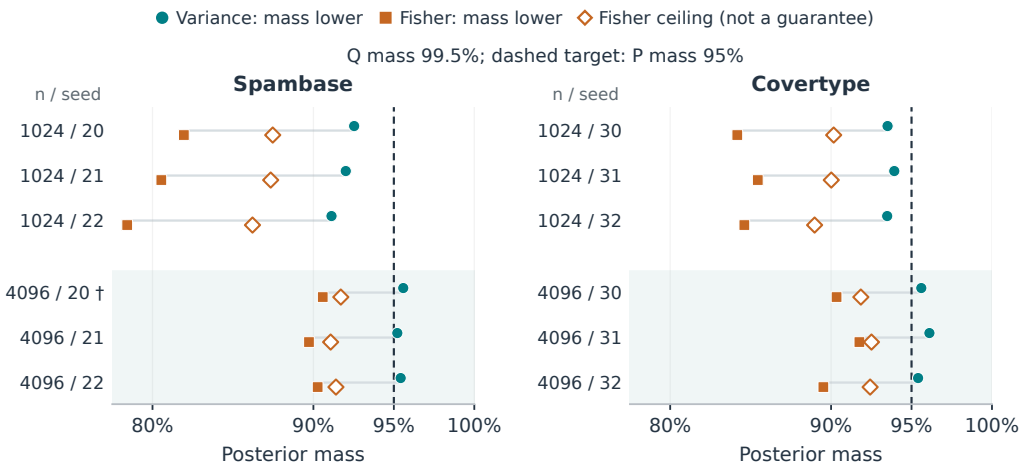


Figure 3: **Posterior-mass lower bounds from KL certificates**, for the fitted Gaussian's 99.5% ellipsoid in every historical $g_{\mathrm{p}} = 4$ VI case. Filled markers are the variance and Fisher lower bounds. Open diamonds are optimistic Fisher ceilings obtained from lower endpoints on its functional; they are not guarantees. The dagger marks a fit whose optimizer returned an "inspect" status rather than a clean convergence flag; it is retained.

mass lower bounds between 95.2% and 96.1%, whereas the Fisher conversions cannot reach 95%. The ellipsoid level and the 95% target were chosen after the KL results and then fixed. These are posterior probabilities under the model, not frequentist coverage.

6.4 Other bounds and limitations

Model-specific controls. We also evaluated a classical Gaussian prior-allocation bound, which applies the geometric Brascamp–Lieb inequality with one optimized allocation scalar, on 24 references at $n = 1024$. In the reported values it is smaller than the degree-resolved bound on the twelve Covertype references and larger on the twelve Spambase references. We treat this comparison as exploratory, because the two implementations reconstruct their inputs differently and exact same-target agreement has not been verified; it was not run on the six $n = 4096$ cases of Figure 3 (Appendix F.5). On eight Poisson references from the quine (Aitkin, 1978) and ships (McCullagh & Nelder, 1983) count data, a classical factor bound is tighter and cheaper (Appendix F.7). On six VI fits selected after favorable results, a Monte Carlo certifier with joint 95% confidence needed 119–187 times the deterministic CPU time to reach tolerance 0.001.

Weak priors and estimated error. At $g_{\mathrm{p}} = 16$, no tested certificate reaches $\mathrm{KL} \le 0.1$; the VI degree-resolved bounds are 0.23–0.29, and these fits also fail the higher-order fitting diagnostic. With unit prior variance, all evaluated bounds are vacuous and the VI fits did not return. Independent diagnostics added after the original study indicate that the moderate-prior failures reflect genuine approximation error rather than a loose certificate. On twelve targets, we refitted VI and estimated $\mathrm{KL}(Q\|P)$ with importance sampling and sequential Monte Carlo (never used as certificate inputs). At $g_{\mathrm{p}} = 4$ the estimated KL is 0.012–0.016 and the certificate of Proposition B.4 is 1.26–1.42 times it; at $g_{\mathrm{p}} = 16$ the estimated KL is 0.12–0.15, already above 0.1, and the ratio is 1.70–1.82. Under unit prior variance the ratio is 17–21.

A genuinely fixed prior. The grids above scale the prior with n. In a separate frozen study on synthetic logistic data with a fixed $N(0, I)$ prior ($d \in \{8, 32\}$, three seeds each, nested $n = 512, \ldots, 32768$), the degree-resolved VI certificate decreased at every sample-size step in all six sequences, from 0.015–0.126 to 0.0003–0.0012, although the certified margin κ fell about fiftyfold. At $d = 8$, $n = 512$ it is 0.015, against 1.04 for matrix Fisher and 0.58 for the scalar variance certificate. Weak global curvature alone is thus not an obstacle once the low-degree terms are measured (Section 3).

7 Related work

Checking approximate posteriors. Most practical checks of Laplace and VI approximations are stochastic. Pareto-smoothed importance sampling diagnoses whether importance weights from Q are reliable (Yao et al., 2018). For data simulated from the model, bidirectional Monte Carlo sandwiches the log marginal likelihood between stochastic lower and upper bounds (Grosse et al., 2015). The χ upper bound (CUBO) upper-bounds the log evidence, but it must itself be estimated by Monte Carlo (Dieng et al., 2017). Huggins et al. (2020) turn such divergence bounds into bounds on Wasserstein distances and on errors in posterior means and variances. Kasprzak et al. (2025) derive computable finite-sample bounds on the error of Laplace approximations in total variation and Wasserstein distances, including for logistic regression. Dehaene (2019) studies log-ratio variance, its cumulant expansion and computable Laplace bounds, including the matrix log-Sobolev bound we use as the Fisher comparator (his Proposition 3), and compares such approximations and bounds empirically on logistic classification; he uses half the variance as an approximation, whereas we prove when the variance certifies. The Gaussian VI stationarity conditions are standard (see, e.g., Opper & Archambeau, 2009), and Katsevich & Rigollet (2024) analyse Gaussian VI accuracy through Hermite expansions; their first-order cancellation underlies our VI corollary. Log-variance objectives are used to fit variational approximations (Richter et al., 2020), were introduced on path space for controlled diffusions (Nüsken & Richter, 2021), and are used to train diffusion-based samplers (Richter & Berner, 2024); our results say when a small log-variance certifies a small KL for a Gaussian approximation. Our bounds are deterministic population inequalities for any fixed Gaussian with global relative-curvature control, not only a Laplace approximation or an exact VI optimum.

Functional inequalities. Logarithmic Sobolev inequalities bound KL by gradient energy (Gross, 1975). The curvature form used as our Fisher comparator goes back to Bakry & Émery (1985); transport and semiconcavity give the same numerator (Gozlan et al., 2011). For concave log ratios, Chen (2026, Theorem 2(1)) obtains the variance bound from convexity of the log moment generating function, sharpening the small-deviation inequality of Paouris & Valettas (2018), and Fernández-Unzueta et al. (2025) extend that convexity to sharp Rényi comparisons in the same $\kappa = 1$ regime; neither covers nonconcave log ratios. Our proof follows the stochastic-control approach to Gaussian functional inequalities (Boué & Dupuis, 1998; Borell, 2000; Lehec, 2013). It uses the version of Hariya & Watanabe (2022) for unbounded functions and the Gaussian covariance representation (Bobkov et al., 2001). For symmetric measures, Fathi et al. (2016) give the scalar degree-two Fisher relaxation; our degree-resolved weights need no symmetry and are matrix-valued. Inverse-Jacobian gradient weights also appear in damped logarithmic Sobolev inequalities on Wiener space (Üstünel, 2001; Üstünel & Zakai, 2007). In the form of Üstünel (2015, Theorem 11.6.1), these are log-Sobolev inequalities for a transformed measure ν, with random weights $(I + \nabla^2 L^{-1} F)^{-1}$ set by the functional F that defines ν; Theorem 3.2 instead bounds $\mathrm{KL}(Q\|P_g)$ itself through Q-expectations with deterministic weights that depend only on K and the Hermite degree. The non-Gaussian extension uses weighted divergence and the transport contraction of Caffarelli (2000) (Villani, 2009; Chewi & Pooladian, 2023). The log-ratio variance is also the squared Fisher–Rao gradient norm of KL; KL/variance comparisons fail without structure (Carrillo et al., 2026) and hold under density-ratio lower bounds (Lu et al., 2023), which our targets need not satisfy.

8 Limitations and conclusion

Limitations. The certificates require global relative-curvature control; a Hessian at the mode is insufficient. On the real-data grids, useful tolerances are reached mainly under strong regularization. The implementation exploits sums of functions of linear projections, needs pairwise row contractions, and uses floating-point arithmetic for the logistic bounds; our KL estimates are diagnostics, not bounds, and model-specific factor bounds can be tighter where they apply.

Conclusion. Relative curvature turns log-ratio variance into a sharp global bound on the error of a Gaussian approximation. Weak curvature is paid only through the stationarity residuals of the fitted Gaussian: at an exact full-covariance VI fit of a smooth log-concave target, $\mathrm{KL}(Q\|P)$ is at most $3/2$ times the log-ratio variance. The certificates are computable without the normalizer, improve as data accumulate under a fixed prior, and stay within about twice the estimated error on the diagnosed targets.

AI-USE STATEMENT

AI assistance was used to help implement the experiments, to aid and polish the writing, for retrieval and discovery (for example, finding related work), to draft sections of the paper, and for proving mathematical claims. The authors take full responsibility for the content of the paper.

REPRODUCIBILITY STATEMENT

The appendices give complete proofs of the stated results, the model and fitting specifications, residual formulas, comparison procedures and the numerical-assurance scope of every reported computation. All reported numerical experiments use CPUs. The historical experiments are exploratory; the prospective Covertype study fixed new rows, settings and reporting rules before fitting. Appendix F distinguishes floating evaluations of analytic bounds from the directed-arithmetic Poisson calculations and lists incomplete and unsuccessful studies.

ETHICS STATEMENT

The empirical study uses public tabular datasets to examine certification of approximate posteriors under specified statistical models. A posterior mass bound is not a guarantee of real-world calibration, predictive accuracy, or suitability for a consequential decision. No deployed decision system is evaluated in this work.

REFERENCES

Murray Aitkin. The analysis of unbalanced cross-classifications. *Journal of the Royal Statistical Society. Series A (General)*, 141(2):195–223, 1978. doi: 10.2307/2344453.

Dominique Bakry and Michel Émery. Diffusions hypercontractives. In Jacques Azéma and Marc Yor (eds.), *Séminaire de Probabilités XIX 1983/84*, volume 1123 of *Lecture Notes in Mathematics*, pp. 177–206. Springer, Berlin, Heidelberg, 1985. doi: 10.1007/BFb0075847.

Jonathan Bennett, Anthony Carbery, Michael Christ, and Terence Tao. The Brascamp–Lieb inequalities: finiteness, structure, and extremals. arXiv:math/0505065v5, 2005. URL https://arxiv.org/abs/math/0505065v5.

Jock Blackard. Covertype. UCI Machine Learning Repository, 1998. URL https://archive.ics.uci.edu/dataset/31/covertype. Dataset.

David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017. doi: 10.1080/01621459.2017.1285773. URL https://arxiv.org/abs/1601.00670.

Sergey G. Bobkov, Friedrich Götze, and Christian Houdré. On Gaussian and Bernoulli covariance representations. *Bernoulli*, 7(3):439–451, 2001. URL https://www-users.cse.umn.edu/~bobko001/papers/2001_Bernoulli_BGH.pdf.

Christer Borell. Diffusion equations and geometric inequalities. *Potential Analysis*, 12(1):49–71, 2000. doi: 10.1023/A:1008641618547.

Michelle Boué and Paul Dupuis. A variational representation for certain functionals of Brownian motion. *The Annals of Probability*, 26(4):1641–1659, 1998. doi: 10.1214/aop/1022855876.

Luis A. Caffarelli. Monotonicity properties of optimal transportation and the FKG and related inequalities. *Communications in Mathematical Physics*, 214(3):547–563, 2000. doi: 10.1007/s002200000257.

José A. Carrillo, Yifan Chen, Daniel Zhengyu Huang, Jiaoyang Huang, and Dongyi Wei. Fisher–Rao gradient flow: Geodesic convexity and functional inequalities, 2026. URL https://arxiv.org/abs/2407.15693v2. arXiv:2407.15693, version 2; first posted in 2024.

Wei-Kuo Chen. A Gaussian convexity for logarithmic moment generating functions with applications in spin glasses. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 62(1), 2026. doi: 10.1214/24-AIHP1527. URL https://arxiv.org/abs/2311.08351v2.

Sinho Chewi and Aram-Alexandre Pooladian. An entropic generalization of Caffarelli's contraction theorem via covariance inequalities. *Comptes Rendus. Mathématique*, 361:1471–1482, 2023. doi: 10.5802/crmath.486. URL https://arxiv.org/abs/2203.04954.

Guillaume P. Dehaene. A deterministic and computable Bernstein–von Mises theorem, 2019. URL https://arxiv.org/abs/1904.02505v2. arXiv:1904.02505, version 2.

Adji Bousso Dieng, Dustin Tran, Rajesh Ranganath, John Paisley, and David Blei. Variational inference via χ upper bound minimization. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/hash/35464c848f410e55a13bb9d78e7fddd0-Abstract.html.

Max Fathi, Emanuel Indrei, and Michel Ledoux. Quantitative logarithmic Sobolev inequalities and stability estimates. *Discrete and Continuous Dynamical Systems*, 36(12):6835–6853, 2016. doi: 10.3934/dcds.2016097. URL https://arxiv.org/abs/1410.6922v1. Cited theorem numbering: arXiv:1410.6922v1.

Maite Fernández-Unzueta, James Melbourne, and Gerardo Palafox-Castillo. On convex functions of Gaussian variables. arXiv:2510.06676, 2025. URL https://arxiv.org/abs/2510.06676.

Nathael Gozlan, Cyril Roberto, and Paul-Marie Samson. A new characterization of Talagrand's transport-entropy inequalities and applications. *The Annals of Probability*, 39(3):857–880, 2011. doi: 10.1214/10-AOP570. URL https://arxiv.org/abs/1104.1303.

Leonard Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975. doi: 10.2307/2373688. URL https://doi.org/10.2307/2373688.

Roger B. Grosse, Zoubin Ghahramani, and Ryan P. Adams. Sandwiching the marginal likelihood using bidirectional Monte Carlo. arXiv:1511.02543, 2015. URL https://arxiv.org/abs/1511.02543.

Yuu Hariya and Sou Watanabe. The Boué–Dupuis formula and the exponential hypercontractivity in the Gaussian space. *Electronic Communications in Probability*, 27, 2022. doi: 10.1214/22-ECP461. URL https://arxiv.org/abs/2110.14852v2. Cited version: arXiv:2110.14852v2.

Mark Hopkins, Erik Reeber, George Forman, and Jaap Suermondt. Spambase. UCI Machine Learning Repository, 1999. URL https://archive.ics.uci.edu/dataset/94/spambase. Dataset.

Jonathan H. Huggins, Mikołaj Kasprzak, Trevor Campbell, and Tamara Broderick. Validated variational inference via practical posterior error bounds. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 1792–1802. PMLR, 2020. URL https://proceedings.mlr.press/v108/huggins20a.html.

Mikołaj J. Kasprzak, Ryan Giordano, and Tamara Broderick. How good is your Laplace approximation of the Bayesian posterior? Finite-sample computable error bounds for a variety of useful divergences. *Journal of Machine Learning Research*, 26(87):1–81, 2025. URL http://jmlr.org/papers/v26/24-0619.html.

Anya Katsevich and Philippe Rigollet. On the approximation accuracy of Gaussian variational inference. *The Annals of Statistics*, 52(4):1384–1409, 2024. doi: 10.1214/24-AOS2393. URL https://arxiv.org/abs/2301.02168v2. Cited version: arXiv:2301.02168v2; section and equation numbers refer to that version.

Béatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338, 2000. doi: 10.1214/aos/1015957395.

Joseph Lehec. Representation formula for the entropy and functional inequalities. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 49(3):885–899, 2013. doi: 10.1214/11-AIHP464.

Yulong Lu, Dejan Slepčev, and Lihan Wang. Birth–death dynamics for sampling: Global convergence, approximations and their asymptotics. *Nonlinearity*, 36(11):5731–5772, 2023. doi: 10.1088/1361-6544/acf988. URL https://arxiv.org/abs/2211.00450.

Peter McCullagh and John A. Nelder. *Generalized Linear Models*. Chapman and Hall, London, 1983.

National Institute of Standards and Technology. NIST Digital Library of Mathematical Functions, section 5.11, 2026. URL https://dlmf.nist.gov/5.11. Version 1.2.8, released September 15, 2026. Accessed September 21, 2026.

Nikolas Nüsken and Lorenz Richter. Solving high-dimensional Hamilton–Jacobi–Bellman PDEs using neural networks: Perspectives from the theory of controlled diffusions and measures on path space. *Partial Differential Equations and Applications*, 2(4):48, 2021. doi: 10.1007/s42985-021-00102-x.

Manfred Opper and Cédric Archambeau. The variational Gaussian approximation revisited. *Neural Computation*, 21(3):786–792, 2009. doi: 10.1162/neco.2008.08-07-592.

Grigoris Paouris and Petros Valettas. A Gaussian small deviation inequality for convex functions. *The Annals of Probability*, 46(3):1441–1454, 2018. doi: 10.1214/17-AOP1206. URL https://arxiv.org/abs/1611.01723.

Lorenz Richter and Julius Berner. Improved sampling via learned diffusions. In *International Conference on Learning Representations*, 2024. URL https://arxiv.org/abs/2307.01198.

Lorenz Richter, Ayman Boustati, Nikolas Nüsken, Francisco J. R. Ruiz, and Ömer Deniz Akyildiz. VarGrad: A low-variance gradient estimator for variational inference. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://arxiv.org/abs/2010.10436.

A. S. Üstünel. Damped logarithmic Sobolev inequality on the Wiener space. In *Stochastic Analysis and Related Topics VII*, pp. 245–249. Birkhäuser Boston, 2001. doi: 10.1007/978-1-4612-0157-1_11.

Ali Süleyman Üstünel. Analysis on Wiener space and applications, 2015. URL https://arxiv.org/abs/1003.1649v2. arXiv:1003.1649, version 2; first posted in 2010.

Ali Süleyman Üstünel and Moshe Zakai. Sufficient conditions for the invertibility of adapted perturbations of identity on the Wiener space. *Probability Theory and Related Fields*, 139(1–2):207–234, 2007. doi: 10.1007/s00440-006-0044-z. Preprint: arXiv:math/0605433, version 3.

William N. Venables and Brian D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002. doi: 10.1007/978-0-387-21706-2.

Cédric Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, Heidelberg, 2009. ISBN 978-3-540-71049-3. doi: 10.1007/978-3-540-71050-9. URL https://link.springer.com/book/10.1007/978-3-540-71050-9.

Yuling Yao, Aki Vehtari, Daniel Simpson, and Andrew Gelman. Yes, but did it work?: Evaluating variational inference. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 5581–5590. PMLR, 2018. URL https://proceedings.mlr.press/v80/yao18a.html.

A THE GAUSSIAN VARIANCE CERTIFICATE

Write $J_Q(g) = \log \mathbb{E}_Q e^g - \mathbb{E}_Q g$ and $P_g = e^g Q / \mathbb{E}_Q e^g$. Whenever these quantities are finite, $J_Q(g) = \mathrm{KL}(Q \| P_g)$; adding a constant to g changes neither the gap nor its variance. We first work with $Q = N(0, I)$.

Theorem A.1 (Gaussian variance comparison)**.** *If $g \in C^2(\mathbb{R}^d) \cap W^{1,2}(Q)$ and $\nabla^2 g \preceq (1-\kappa) I$ everywhere, where $0 < \kappa \leq 1$, then P_g is well-defined, both relative entropies between Q and P_g are finite, and*

$$J_Q(g) \leq \frac{\mathrm{Var}_Q(g)}{2\kappa}. \tag{9}$$

The coefficient is optimal for each fixed κ, already in dimension one. Here $W^{1,2}(Q)$ requires $g, \nabla g \in L^2(Q)$.

A.1 BROWNIAN PROOF, INCLUDING UNBOUNDED LOG RATIOS

Proof of the inequality. Taylor's upper bound gives

$$\begin{aligned} g(z) &\leq g(0) + \nabla g(0)^\top z + \frac{1-\kappa}{2} \|z\|^2, \\ 0 < \mathbb{E}_Q e^g &\leq e^{g(0)} \kappa^{-d/2} e^{\|\nabla g(0)\|^2/(2\kappa)} < \infty. \end{aligned} \tag{10}$$

This also gives Gaussian upper tails for P_g. The positive part of g has a quadratic upper bound, and $(-g)e^g \leq e^{-1}$ on $\{g < 0\}$, so $\mathbb{E}_{P_g}|g| < \infty$. Together with $g \in L^2(Q)$, this proves finiteness of both entropies.

Let W be standard Brownian motion on $[0,1]$ with its augmented natural filtration $(\mathcal{F}_t)$, and let $\mathcal{A}$ be the progressively measurable drifts with $\mathbb{E}\int_0^1 \|u_t\|^2 \, \mathrm{d}t < \infty$. Since $g \in L^1(Q)$ and $\mathbb{E}_Q e^g < \infty$, the integrable terminal-value Boué–Dupuis representation (Hariya & Watanabe, 2022, Proposition 3.1) gives

$$\log \mathbb{E} e^{g(W_1)} = \sup_{u \in \mathcal{A}} \mathbb{E}\left[g\left(W_1 + \int_0^1 u_t \, \mathrm{d}t \right) - \frac{1}{2} \int_0^1 \|u_t\|^2 \, \mathrm{d}t \right]. \tag{11}$$

For every admissible drift the shifted log ratio has integrable positive part: its quadratic upper envelope is integrable because $W_1 + \int_0^1 u_t \, \mathrm{d}t \in L^2$. Its negative part may have infinite expectation, in which case the objective is $-\infty$ (Hariya & Watanabe, 2022, Remark 1.1(1)).

Put $h = \int_0^1 u_t \, \mathrm{d}t$ and $m_t = \mathbb{E}[\nabla g(W_1) \mid \mathcal{F}_t]$, with its continuous square-integrable martingale version. Semiconcavity and $\|h\|^2 \leq \int_0^1 \|u_t\|^2 \, \mathrm{d}t$ imply

$$\begin{aligned} &\mathbb{E}\left[g(W_1 + h) - g(W_1) - \frac{1}{2} \int_0^1 \|u_t\|^2 \, \mathrm{d}t \right] \\ &\qquad \leq \mathbb{E} \int_0^1 \left[m_t^\top u_t - \frac{\kappa}{2} \|u_t\|^2 \right] \mathrm{d}t \leq \frac{1}{2\kappa} \mathbb{E} \int_0^1 \|m_t\|^2 \, \mathrm{d}t. \end{aligned} \tag{12}$$

The first inequality uses $1 - \kappa \geq 0$. Adapted conditioning replaces $\nabla g(W_1)$ by m_t; Cauchy–Schwarz justifies conditioning and Fubini from the two square-integrability assumptions. The second inequality is pointwise completion of the square.

The terminal Clark–Ocone representation and Itô isometry give

$$g(W_1) - \mathbb{E}_Q g = \int_0^1 m_t^\top \, \mathrm{d}W_t, \qquad \mathrm{Var}_Q(g) = \mathbb{E} \int_0^1 \|m_t\|^2 \, \mathrm{d}t. \tag{13}$$

For smooth compactly supported g, the backward heat equation and Itô's formula prove this identity. Approximation in $W^{1,2}(Q)$, contraction of conditional expectation in L^2, and Itô isometry pass both the integrands and stochastic integrals to the limit. Taking the supremum in (12) proves (9); no optimizing drift is needed. □

We will also use the covariance form of this argument. Conditionally independent copies of W_1 given $\mathcal{F}_t$ are standard Gaussian with cross-covariance tI. Thus, for a constant matrix B, independent $X, Y \sim Q$, and $f = \nabla g$,

$$\mathbb{E}[m_t^\top B m_t] = \mathbb{E}\Big[f(X)^\top B f\Big(tX + \sqrt{1-t^2}Y\Big)\Big]. \tag{14}$$

For $B = I$, integration in t is the Gaussian covariance formula of Bobkov et al. (2001, equation (1.1)). The same L^2 approximation justifies it in the present class.

Sharpness and whitening. For $g_b(z) = (1-\kappa)z^2/2 + bz$, Gaussian integration gives

$$J_Q(g_b) = \frac{-\log\kappa - (1-\kappa)}{2} + \frac{b^2}{2\kappa}, \qquad \mathrm{Var}_Q(g_b) = \frac{(1-\kappa)^2}{2} + b^2. \tag{15}$$

Their ratio tends to $1/(2\kappa)$ as $|b| \to \infty$; affine log ratios attain equality when $\kappa = 1$. For Theorem 3.1, set $\widehat{g}(z) = g(\mu + \Sigma^{1/2}z)$. Then $\nabla\widehat{g} = \Sigma^{1/2}\nabla g$ and

$$\nabla^2\widehat{g} = I + \Sigma^{1/2}\nabla^2(\log\widetilde{p})\Sigma^{1/2} \preceq (1-\kappa)I.$$

The Sobolev assumptions transfer exactly. Apply Theorem A.1 and invariance of KL and variance under this invertible change of variables. Since $\mathbb{E}_Q e^g = \int \widetilde{p}$, normalizability follows as well.

An equivalent Bernstein form. Let $F \in C^2 \cap W^{1,2}(Q)$ satisfy $\nabla^2 F \preceq \rho I$ with $\rho \geq 0$, and let $\lambda \geq 0$ with $\lambda\rho < 1$. Applying Theorem A.1 to $g = \lambda F$, whose curvature margin is $\kappa = 1 - \lambda\rho$, gives

$$\log \mathbb{E}_Q e^{\lambda(F - \mathbb{E}_Q F)} \leq \frac{\lambda^2 \,\mathrm{Var}_Q(F)}{2(1-\lambda\rho)}.$$

This is a one-sided Bernstein bound whose variance factor is the exact variance; conversely, the bound for all such F and λ returns Theorem A.1. For $F = \sum_i a_i(Z_i^2 - 1)$ with $a_i \geq 0$, $\nabla^2 F \preceq 2\max_i a_i\, I$ and $\mathrm{Var}_Q(F) = 2\sum_i a_i^2$, and the bound is the chi-square log-Laplace bound in the proof of Laurent & Massart (2000, Lemma 1).

B THE DEGREE-DEPENDENT MATRIX CERTIFICATE

Throughout this section $Q = N(0, I)$, $g \in C^2 \cap W^{1,2}(Q)$, and g_k is the projection onto Gaussian chaos of total degree k. For a constant symmetric matrix $K \succ 0$, define

$$\begin{aligned} R_t &= [K + (I-K)t]^{-1}, \qquad W_k(K) = \int_0^1 t^{k-1} R_t^2 \,\mathrm{d}t, \\ \mathcal{E}_K(g) &= \sum_{k\geq 1} \mathbb{E}_Q[(\nabla g_k)^\top W_k(K) \nabla g_k]. \end{aligned} \tag{16}$$

Matrix functions act on the eigenvalues of K; in particular,

$$W_1(K) = K^{-1}, \qquad W_2(\lambda) = \frac{\lambda - 1 - \log\lambda}{(\lambda-1)^2}, \qquad W_2(1) = \frac{1}{2}. \tag{17}$$

The differentiated chaos expansion converges in $L^2(Q)$, and $0 \prec W_k(K) \preceq K^{-1}$, so the energy is finite. For comparison define $\mathcal{D}_A(g) = \sum_{k\geq 1} k^{-1}\mathbb{E}_Q[(\nabla g_k)^\top A^{-1}\nabla g_k]$, for $A \succ 0$.

Theorem B.1 (Quadratic-control certificate)**.** *If $\nabla^2 g \preceq I - K$ everywhere, with $K \succ 0$, then*

$$J_Q(g) \leq \tfrac{1}{2}\mathcal{E}_K(g). \tag{18}$$

Equality holds for every $g(z) = z^\top(I-K)z/2 + b^\top z + c$, so the coefficient $1/2$ is optimal for each fixed K. Let K_{cap} replace each eigenvalue λ_i of K by $\min(1, \lambda_i)$, and set $\kappa_* = \min(1, \lambda_{\min}(K))$. Then

$$\mathcal{E}_K(g) \leq \min\big\{\mathbb{E}_Q[(\nabla g)^\top K^{-1}\nabla g], \mathcal{D}_{K_{\mathrm{cap}}}(g)\big\} \leq \frac{\mathrm{Var}_Q(g)}{\kappa_*}. \tag{19}$$

The last inequality uses the second argument of the minimum.

There is no upper restriction on K. Gaussian equality requires matching curvature and does not assert separate optimality of every higher-degree weight. At $K = I$, $W_k = I/k$, recovering the concave endpoint (Chen, 2026). For even log ratios and scalar curvature, the degree-two relaxation $J_Q(g) \le W_2(\kappa)\mathbb{E}_Q \|\nabla g\|^2 /2$ is already in Fathi et al. (2016, Theorem 12 in the cited preprint).

Proof. Put $H = I - K$ and use the Brownian notation above. The quadratic upper envelope and $K \succ 0$ ensure normalizability and justify (11), including its extended objectives. For $h_t = \int_0^t u_s \, \mathrm{d}s$, semiconcavity gives $J_Q(g) \le \sup_{u \in \mathcal{A}} L(u)$, where

$$L(u) = \mathbb{E}\left[m_1^\top h_1 + \tfrac{1}{2} h_1^\top H h_1 - \tfrac{1}{2} \int_0^1 \|u_t\|^2 \, \mathrm{d}t \right]. \tag{20}$$

This step permits indefinite H. Set $C_t = (1-t)R_t$, so $R_t' = -R_t H R_t$, $C_t' = -R_t^2$, and define

$$F(t,h,m) = \tfrac{1}{2} h^\top H R_t h + h^\top R_t m + \tfrac{1}{2} m^\top C_t m, \qquad p_t = R_t(Hh_t + m_t).$$

Itô's formula, initially localized, yields

$$\begin{aligned} \mathrm{d}F(t,h_t,m_t) = {} & \tfrac{1}{2}\big(\|u_t\|^2 - \|u_t - p_t\|^2\big)\,\mathrm{d}t + (R_t h_t + C_t m_t)^\top \,\mathrm{d}m_t \\ & + \tfrac{1}{2}\operatorname{tr}(C_t \,\mathrm{d}\langle m \rangle_t). \end{aligned} \tag{21}$$

The endpoint values are $F(0,0,m_0) = m_0^\top K^{-1} m_0/2$ and $F(1,h_1,m_1) = h_1^\top H h_1/2 + h_1^\top m_1$.

Here taking expectations requires only second moments. Doob's inequality and finite control energy give $\mathbb{E}\sup_t \|m_t\|^2 < \infty$, $\mathbb{E}\sup_t \|h_t\|^2 < \infty$, and $\mathbb{E}\operatorname{tr}\langle m \rangle_1 < \infty$. Since R_t, C_t are bounded, the stochastic integral $N = \int (R_t h_t + C_t m_t)^\top \,\mathrm{d}m_t$ satisfies

$$\mathbb{E}\sqrt{\langle N \rangle_1} \le \big(\mathbb{E}\sup_t \|R_t h_t + C_t m_t\|^2\big)^{1/2} \big(\mathbb{E}\operatorname{tr}\langle m \rangle_1\big)^{1/2} < \infty.$$

Burkholder–Davis–Gundy gives an integrable running maximum, hence uniform integrability and zero mean. Applying the same argument to $m_t^\top C_t m_t$, whose terminal value is zero, gives

$$m_0^\top K^{-1} m_0 + \mathbb{E}\int_0^1 \operatorname{tr}(C_t \,\mathrm{d}\langle m \rangle_t) = \mathbb{E}\int_0^1 m_t^\top R_t^2 m_t \,\mathrm{d}t.$$

Thus (21) proves the exact completion identity

$$L(u) = \tfrac{1}{2}\mathbb{E}\int_0^1 m_t^\top R_t^2 m_t \,\mathrm{d}t - \tfrac{1}{2}\mathbb{E}\int_0^1 \|u_t - R_t(Hh_t + m_t)\|^2 \,\mathrm{d}t. \tag{22}$$

The feedback $\dot{h}_t = R_t(Hh_t + m_t)$, $h_0 = 0$, has a unique adapted solution of finite expected energy, since its coefficients are bounded. Equivalently, $u_t = K^{-1} m_0 + \int_0^t R_s \,\mathrm{d}m_s$. It attains the relaxed value; this does not assert that it maximizes the original objective for a nonquadratic log ratio.

In (14), ∇g_k has Hermite degree $k-1$, giving the factor t^{k-1}; different degrees are orthogonal. Consequently $\mathbb{E}\int_0^1 m_t^\top R_t^2 m_t \,\mathrm{d}t = \mathcal{E}_K(g)$. Boundedness of R_t and L^2 convergence pass this identity from finite chaos sums to g, proving the bound. For a matching quadratic the semiconcavity step is exact, and

$$\tfrac{1}{2}\mathcal{E}_K(g) = \tfrac{1}{2} b^\top K^{-1} b + \tfrac{1}{2}[-\log\det K - \operatorname{tr}(I-K)] = J_Q(g). \tag{23}$$

For $K \ne I$ the quadratic part is nonconstant; for $K = I$ choose $b \ne 0$. This proves sharpness.

Finally, $W_k(K) \preceq W_1(K) = K^{-1}$. For $k \ge 2$, integration by parts using $C_t' = -R_t^2$ gives

$$W_k(K) = (k-1)\int_0^1 t^{k-2}(1-t)[tI + (1-t)K]^{-1} \,\mathrm{d}t. \tag{24}$$

An eigenvalue denominator is at least $\min(1,\lambda)$, hence $W_k(\lambda) \le 1/[k\min(1,\lambda)]$. Together with $W_1(\lambda) = 1/\lambda$ and $\mathbb{E}_Q\|\nabla g_k\|^2 = k\operatorname{Var}_Q(g_k)$, this proves (19). The positive inverse representation also makes W_k order-reversing on positive definite matrices. For $k \ge 3$, the original denominator is at least t, giving $W_k(K) \preceq I/(k-2)$. □

B.1 Low-degree weights, exact Gaussian VI, and a quartic example

Proof of Corollary 3.3. Apply Theorem B.1 with $K = \kappa I$. The weights are scalar: $W_1(\kappa) = 1/\kappa$ and $W_2(\kappa)$ is given by (17). For $k \geq 3$, $\kappa+(1-\kappa)t \geq t$ on $[0,1]$, so $W_k(\kappa) \leq \int_0^1 t^{k-3}\,\mathrm{d}t = 1/(k-2)$. Since $\mathbb{E}_Q \|\nabla g_k\|^2 = kV_k$, degree k contributes at most $kW_k(\kappa)V_k/2$. This is $V_1/(2\kappa)$ for $k = 1$ and $W_2(\kappa)V_2$ for $k = 2$. For $k \geq 3$ it is at most $kV_k/[2(k-2)] \leq \frac{3}{2}V_k$. If g is even, its odd-degree components vanish. If also $g_2 = 0$, only degrees $k \geq 4$ remain, and $k/[2(k-2)] \leq 1$ gives $J_Q(g) \leq V_Q(g)$. □

Lemma B.2 (Low-degree components and the Gaussian VI objective)**.** *Let $\log \widetilde{p} \in C^2(\mathbb{R}^d)$ have first and second derivatives of at most polynomial growth. Let $\mathcal{F}(m, M) = -\mathbb{E} \log \widetilde{p}(m + MZ) - \log|\det M|$, $Z \sim N(0, I)$, be the reverse-KL objective up to a constant, over $m \in \mathbb{R}^d$ and invertible M. For invertible L, let $Q = N(\mu, LL^\top)$ and $g(z) = \log \widetilde{p}(\mu + Lz) + \|z\|^2/2$. Then $g \in C^2 \cap W^{1,2}(Q)$, $\nabla^2 g \in L^1(Q)$, and the Hermite components of degrees one and two are*

$$g_1(z) = a^\top z, \qquad g_2(z) = \tfrac{1}{2}(z^\top B z - \operatorname{tr} B),$$

where $a = \mathbb{E}_Q \nabla g = -L^\top \nabla_m \mathcal{F}(\mu, L)$ and $B = \mathbb{E}_Q \nabla^2 g = -L^\top \nabla_M \mathcal{F}(\mu, L)$. In particular, $g_1 = g_2 = 0$ when (μ, L) is a stationary point.

Proof. Polynomial growth makes g, its first two derivatives, the integrands and their parameter derivatives dominated by a polynomial in Z, locally uniformly in (m, M). So $g \in W^{1,2}(Q)$, $\nabla^2 g \in L^1(Q)$, and one may differentiate under the expectation and integrate by parts against the Gaussian. Writing $X = \mu + LZ$, $\nabla_m \mathcal{F} = -\mathbb{E}\nabla \log \widetilde{p}(X)$ and $\nabla_M \mathcal{F} = -\mathbb{E}[\nabla \log \widetilde{p}(X) Z^\top] - L^{-\top}$ at (μ, L). Gaussian integration by parts gives $\mathbb{E}[\nabla \log \widetilde{p}(X) Z^\top] = \mathbb{E}[\nabla^2 \log \widetilde{p}(X)]L$. In whitened coordinates $\nabla g = L^\top \nabla \log \widetilde{p} + z$ and $\nabla^2 g = L^\top \nabla^2 \log \widetilde{p}\, L + I$. Hence $\mathbb{E}_Q \nabla g = -L^\top \nabla_m \mathcal{F}$ and $\mathbb{E}_Q \nabla^2 g = L^\top \mathbb{E}[\nabla^2 \log \widetilde{p}(X)]L + I = -L^\top \nabla_M \mathcal{F}$. Integration by parts again gives $\mathbb{E}_Q[g(Z) Z_i] = \mathbb{E}_Q \partial_i g$ and $\mathbb{E}_Q[g(Z)(Z_i Z_j - \delta_{ij})] = \mathbb{E}_Q \partial_{ij} g$. These are the degree-one and degree-two Hermite coefficients of g; summing them against the normalized Hermite basis gives the stated g_1 and g_2. At a stationary point both gradients vanish. The resulting cancellation is the one used by Katsevich & Rigollet (2024, Section 4.2); the stationarity conditions are standard (see, e.g., Opper & Archambeau, 2009). □

Proof of Corollary 3.4. By Lemma B.2, $g \in C^2 \cap W^{1,2}(Q)$ and $g_1 = g_2 = 0$; log-concavity of $\widetilde{p}$ gives $\nabla^2 g \preceq I$. Fix $0 < \theta < 1$. Then $\theta g \in C^2 \cap W^{1,2}(Q)$, $\nabla^2(\theta g) \preceq (1 - (1-\theta))I$, the degree-one and degree-two components of θg vanish, and $\operatorname{Var}_Q(\theta g) = \theta^2 \operatorname{Var}_Q(g)$. Corollary 3.3 with $\kappa = 1 - \theta$, and its even case proved above, give

$$\log \mathbb{E}_Q e^{\theta g} \leq \theta \mathbb{E}_Q g + c\,\theta^2 \operatorname{Var}_Q(g), \qquad c = \tfrac{3}{2}, \text{ or } c = 1 \text{ if } g \text{ is even.}$$

As $\theta \uparrow 1$, $e^{\theta g} \to e^g$ pointwise, and Fatou's lemma gives $\mathbb{E}_Q e^g \leq \liminf_{\theta\uparrow 1} \mathbb{E}_Q e^{\theta g} \leq \exp(\mathbb{E}_Q g + c \operatorname{Var}_Q(g)) < \infty$. Since $\widetilde{p}(\mu + Lz) = (2\pi)^{d/2} e^{g(z)} \gamma_d(z)$, where γ_d is the standard Gaussian density, $\int \widetilde{p} = |\det L|(2\pi)^{d/2} \mathbb{E}_Q e^g \in (0, \infty)$. Hence P is well defined, $\mathrm{KL}(Q\|P) = J_Q(g)$ by (2), and $J_Q(g) \leq c \operatorname{Var}_Q(g)$. □

Exact stationarity is a property of the population objective. A numerically fitted approximation satisfies it only approximately. The computations of Section 6 therefore keep the measured degree-one and degree-two terms, with their K^{-1} and $W_2(K)$ weights, and charge polynomial residuals through (25). Proposition B.4 gives an alternative in which the weak-curvature weights act only on the exactly computed stationarity residuals a and B.

Proposition B.3 (An exactly fitted quartic)**.** *Let $Q = N(0, 1)$, $0 < \kappa < 1$, $H_4(z) = z^4 - 6z^2 + 3$, and $g_\kappa = -(1-\kappa)H_4/12$. Then:*

1. *$g_\kappa''(z) = (1-\kappa)(1-z^2) \leq 1 - \kappa$, and $P_\kappa \propto \exp[-\kappa z^2/2 - (1-\kappa)z^4/12]$.*
2. *g_κ is pure Hermite degree four, with $V_Q(g_\kappa) = (1-\kappa)^2/6$. The scalar certificate equals $(1-\kappa)^2/(12\kappa)$. The degree-resolved certificate equals $(1-\kappa)^2 W_4(\kappa)/3 \leq (1-\kappa)^2/6$, and tends to $1/6$ as $\kappa \to 0$.*

3. $J_Q(g_\kappa) \to J_0 = \log[12^{1/4}\Gamma(1/4)/(2\sqrt{2\pi})] - \frac{1}{4}$, *which is approximately* 0.0472 *and strictly positive.*

4. Q *is the unique minimizer of* $\mathrm{KL}(N(m, s^2)\|P_\kappa)$ *over* $m \in \mathbb{R}$, $s > 0$.

Proof. (1) This follows by differentiation, and by adding g_κ to $-z^2/2$. (2) $\mathrm{Var}_Q(H_4) = 4! = 24$ gives the variance. The scalar certificate is $V_Q/(2\kappa)$. The degree-four energy is $\mathbb{E}_Q(g'_\kappa)^2 = 4V_Q(g_\kappa)$, and Theorem B.1 with $K = \kappa$ contributes $\frac{1}{2}W_4(\kappa) \cdot 4V_Q(g_\kappa)$. Since $\kappa + (1-\kappa)t \geq t$, $W_4(\kappa) \leq \int_0^1 t\,\mathrm{d}t = \frac{1}{2}$. By monotone convergence, $W_4(\kappa) \to \frac{1}{2}$ as $\kappa \to 0$. (3) $\mathbb{E}_Q g_\kappa = 0$, and $\mathbb{E}_Q e^{g_\kappa} = e^{-(1-\kappa)/4}(2\pi)^{-1/2}\int e^{-\kappa z^2/2-(1-\kappa)z^4/12}\,\mathrm{d}z$. By dominated convergence this tends to $e^{-1/4}(2\pi)^{-1/2}\int e^{-z^4/12}\,\mathrm{d}z$. Moreover $\int_{\mathbb{R}} e^{-z^4/12}\,\mathrm{d}z = 2 \cdot 12^{1/4}\Gamma(5/4) = 12^{1/4}\Gamma(1/4)/2$. Positivity holds because J_0 is the KL divergence from $N(0,1)$ to the non-Gaussian law with density proportional to $e^{-z^4/12}$. (4) Up to a constant, the objective is $\mathbb{E}U_\kappa(m + sZ) - \log s$, where $U_\kappa(x) = \kappa x^2/2 + (1-\kappa)x^4/12$ is strictly convex. Distinct (m, s) give random variables $m + sZ$ that differ almost surely, so the first term is strictly convex in (m, s); adding the convex $-\log s$, any minimizer is unique, and since the objective is convex and differentiable, every stationary point is the minimizer. The objective is even in m, so $m = 0$. At $m = 0$ the objective is $\kappa s^2/2 + (1-\kappa)s^4/4 - \log s$. Its stationarity equation, $\kappa s^2 + (1-\kappa)s^4 = 1$, has the unique positive root $s = 1$. □

The margin-dependent theorems do not apply at $\kappa = 0$, but Corollary 3.4 does. The endpoint target $P_0 \propto e^{-z^4/12}$ is log-concave, and the argument of item (4) with $\kappa = 0$ shows that Q is its exact Gaussian VI fit (the stationarity equation becomes $s^4 = 1$). Since $g_0 = -H_4/12$ is even, the corollary gives $J_Q(g_0) \leq V_Q(g_0) = 1/6$, the limit of the degree-resolved certificates, against the true value $J_0 \approx 0.047$.

B.2 Finite-polynomial computation and residual bounds

Under Theorem B.1, let P be a polynomial with $\|g - P\|_{L^2(Q)} \leq r_v$ and $\left\|K^{-1/2}\nabla(g - P)\right\|_{L^2(Q)} \leq r_f$. The comparisons in (19) are spectral inequalities valid for any Sobolev function; the residual needs no curvature condition. Since $\mathcal{E}_K^{1/2}$ is a seminorm and $\mathrm{Var}_Q(g - P) \leq r_v^2$, its triangle inequality gives

$$J_Q(g) \leq \frac{1}{2}\left[\sqrt{\mathcal{E}_K(P)} + \min\{r_v/\sqrt{\kappa_*}, r_f\}\right]^2. \tag{25}$$

The coefficients need not be exact projection coefficients, but the residual must refer to the same polynomial whose energy is evaluated. Writing P_k for its degree-k Hermite component, the energy is a finite sum of Gaussian polynomial moments: each degree's gradient covariance contracts with $W_k(K)$.

If only $e_k = \mathbb{E}_Q\|\nabla P_k\|^2 = k\,\mathrm{Var}_Q(P_k)$ and $F_k = \mathbb{E}_Q[(\nabla P_k)^\top K^{-1}\nabla P_k]$ are available, one can instead use the following scalar upper bound. For $k \geq 2$, $f_k(y) = W_k(1/y)$ is concave: (24) expresses it as a positive integral of $y/[ty + (1-t)]$. For $e_k > 0$, Jensen's inequality with the normalized gradient energies in the eigenbasis of K gives

$$\mathbb{E}_Q[(\nabla P_k)^\top W_k(K)\nabla P_k] \leq e_k f_k(F_k/e_k) = (k-1)\int_0^1 \frac{t^{k-2}(1-t)e_k F_k}{tF_k + (1-t)e_k}\,\mathrm{d}t. \tag{26}$$

Degree one contributes F_1; a component with $e_k = 0$ contributes zero. A valid upper bound on F_k can be used for a small component without division by e_k. The scalar relaxation can lose exactness for anisotropic quadratics. These are real-arithmetic bounds; numerical evaluation must enclose the polynomial, residual, and spectral terms.

The weak-curvature weights K^{-1} and $W_2(K)$ need not touch the polynomial residual. Write $\mathcal{E}_{K,\geq 3}$ for the sum in the definition of $\mathcal{E}_K$ restricted to degrees $k \geq 3$.

Proposition B.4 (Low-degree terms computed on g)**.** *Under the assumptions of Theorem B.1, assume also* $\nabla^2 g \in L^1(Q)$*, and let* $a = \mathbb{E}_Q\nabla g$ *and* $B = \mathbb{E}_Q\nabla^2 g$*. Let* P *be a polynomial with*

$\|g - P\|_{L^2(Q)} \le r_v$ *and* $\|K^{-1/2}\nabla(g-P)\|_{L^2(Q)} \le r_f$*, and set* $c_* = \min\{\sqrt{3}, \kappa_*^{-1/2}\}$*. Then*

$$J_Q(g) \le \frac{1}{2}a^\top K^{-1}a + \frac{1}{2}\operatorname{tr}\big(BW_2(K)B\big) + \frac{1}{2}\left[\sqrt{\mathcal{E}_{K,\ge 3}(P)} + \min\{c_* r_v, r_f\}\right]^2.$$

Proof. As in the proof of Lemma B.2, Gaussian integration by parts gives $g_1(z) = a^\top z$ and $g_2(z) = \frac{1}{2}(z^\top Bz - \operatorname{tr} B)$, so the degree-one and degree-two terms of $\mathcal{E}_K(g)$ equal $a^\top K^{-1}a$ and $\operatorname{tr}(BW_2(K)B)$. For $k \ge 3$, $K + (I-K)t \succeq tI$ gives $W_k(K) \preceq I/(k-2)$, and $W_k(K) \preceq W_1(K) = K^{-1}$; the proof of Theorem B.1 also gives $W_k(K) \preceq I/(k\kappa_*)$. Hence, for $r = g - P$, $\mathcal{E}_{K,\ge 3}(r) \le \sum_{k\ge 3} \min\{\frac{k}{k-2}, \kappa_*^{-1}\}\operatorname{Var}_Q(r_k) \le c_*^2 r_v^2$ and $\mathcal{E}_{K,\ge 3}(r) \le \mathbb{E}_Q[\nabla r^\top K^{-1}\nabla r] \le r_f^2$. Since $\mathcal{E}_{K,\ge 3}^{1/2}$ is a seminorm and $(g-P)_k = g_k - P_k$, the triangle inequality bounds $\mathcal{E}_{K,\ge 3}(g)^{1/2}$. Theorem B.1 completes the proof. □

For a generalized linear model, a and B are sums of one-dimensional Gaussian integrals, one per observation, and can be enclosed directly. The curvature margin then enters only through the stationarity residuals of the fitted Gaussian, which a converged VI solver makes small, while the polynomial residual is charged with the constant $c_* \le \sqrt{3}$, which is curvature-free. The reported logistic evaluations use (25).

C References with two-sided curvature control

Theorem C.1 (General reference comparison)**.** *Let Q have density e^{-U}/Z_Q, where $U \in C^2(\mathbb{R}^d)$ and*

$$\alpha I \preceq \nabla^2 U \preceq \beta I, \qquad 0 < \alpha \le \beta < \infty. \tag{27}$$

If $g \in C^2 \cap W^{1,2}(Q)$ and $\nabla^2 g \preceq \rho I$ everywhere, with $0 \le \rho < \alpha$, then P_g is well-defined, both relative entropies are finite, and

$$J_Q(g) \le \frac{\beta}{2(\alpha-\rho)}\operatorname{Var}_Q(g). \tag{28}$$

For each fixed α, β, ρ, the coefficient is optimal uniformly over this source and log-ratio class, already in dimension one.

C.1 Weighted divergence and the transport bound

Lemma C.2 (Weighted divergence identity)**.** *Under* (27)*, set $\delta_Q u = \nabla U^\top u - \operatorname{div} u$. For a smooth compactly supported vector field u,*

$$\mathbb{E}_Q \delta_Q u = 0, \qquad \mathbb{E}_Q[\nabla f^\top u] = \mathbb{E}_Q[f\delta_Q u], \qquad f \in W^{1,2}(Q), \tag{29}$$

$$\mathbb{E}_Q(\delta_Q u)^2 = \mathbb{E}_Q[u^\top(\nabla^2 U)u] + \mathbb{E}_Q \operatorname{tr}((Du)^2). \tag{30}$$

The identities extend to globally Lipschitz u of at most linear growth, with weak Jacobian Du. For symmetric Du, the trace is its squared Frobenius norm.

Proof. Integration by parts gives $\mathbb{E}_Q \partial_i v = \mathbb{E}_Q[v\partial_i U]$ and the adjoint identity. Writing $U_i = \partial_i U$, a second application gives

$$\begin{aligned}\mathbb{E}_Q(\delta_Q u)^2 &= \mathbb{E}_Q \sum_{i,j} u_i U_{ij} u_j + \mathbb{E}_Q \sum_{i,j} u_i(U_j \partial_i u_j - \partial_j \partial_i u_j) \\ &= \mathbb{E}_Q[u^\top(\nabla^2 U)u] + \mathbb{E}_Q \sum_{i,j}(\partial_j u_i)(\partial_i u_j).\end{aligned}$$

For the extension, lower curvature gives Gaussian upper tails and upper curvature makes ∇U grow at most linearly. Thus $u, Du, \delta_Q u$ have the required square moments. Multiply u by smooth cutoffs on growing balls, then mollify. The fields, Jacobians, and weighted divergences converge in their respective $L^2(Q)$ spaces; the cutoff errors vanish by the Gaussian tails. Pass to the limit in both identities, and approximate f in $W^{1,2}(Q)$ for adjointness. □

Lemma C.3 (Jacobian remainder). *For $r \geq 1$ and $-1 < a \leq r-1$, $a^2 \leq 2r[a - \log(1+a)]$.*

Proof. For $a \geq 0$, $a - \log(1+a) = \int_0^a t/(1+t)\,\mathrm{d}t \geq a^2/(2r)$. For $a = -b \in (-1, 0)$, the same expression is $\int_0^b t/(1-t)\,\mathrm{d}t \geq b^2/2$. □

Proof of Theorem C.1, inequality. Put $\kappa = \alpha - \rho$, $v = \mathrm{Var}_Q(g)$, and $r = \sqrt{\beta/\kappa} \geq 1$. The quadratic bounds on U and the upper quadratic envelope for g make $U - g$ grow at least quadratically with positive coefficient κ. Thus P_g exists and has all polynomial moments. Its log ratio is integrable: the positive part is bounded by a quadratic and $(-g)e^g \leq e^{-1}$ on $\{g < 0\}$. Both entropies are finite.

Let $T : Q \to P_g$ be the Brenier map. The source upper curvature β and target lower curvature κ give the classical contraction comparison (Chewi & Pooladian, 2023), $0 \prec DT \preceq rI$ almost everywhere. Its Jacobian is symmetric, as T is a convex gradient, and positive definite by change of variables for the positive densities. Set

$$u = T - \mathrm{id}, \quad A = Du = DT - I, \quad x_0 = \tfrac{1}{2}\mathbb{E}_Q \left\|u\right\|^2, \quad y_0 = \mathbb{E}_Q \,\mathrm{tr}[A - \log(I + A)].$$

The Lipschitz map has linear growth. Also $\log\det DT(X) = \log q(X) - \log p_g(T(X))$ is integrable, because $U(X), U(T(X)), g(T(X))$ are integrable. Hence x_0, y_0 are finite and nonnegative, without a uniform lower bound on DT. Change of variables and weighted adjointness give

$$D := \mathrm{KL}(P_g \| Q) = \mathbb{E}_Q[U(T) - U - \nabla U^\top u] + y_0 \geq \alpha x_0 + y_0. \tag{31}$$

The eigenvalues of A lie in $(-1, r-1]$. The two lemmas imply

$$\mathbb{E}_Q(\delta_Q u)^2 = \mathbb{E}_Q[u^\top(\nabla^2 U)u] + \mathbb{E}_Q \,\mathrm{tr}(A^2) \leq 2\beta x_0 + 2r y_0. \tag{32}$$

Semiconcavity, adjointness, and Cauchy–Schwarz therefore give

$$\begin{aligned} J_Q(g) + D = \mathbb{E}_Q[g(T) - g] &\leq \mathbb{E}_Q[(g - \mathbb{E}_Q g)\delta_Q u] + \rho x_0 \\ &\leq \sqrt{2v(\beta x_0 + r y_0)} + \rho x_0. \end{aligned}$$

For $s = \beta x_0 + r y_0$, $\kappa x_0 + y_0 \geq (\kappa/\beta)s$, since $\kappa r/\beta = \sqrt{\kappa/\beta} \leq 1$. Consequently

$$J_Q(g) \leq \sqrt{2vs} - (\kappa/\beta)s \leq \frac{\beta v}{2\kappa}.$$

All adjoint identities apply by Lemma C.2: u is Lipschitz, $\delta_Q u \in L^2(Q)$, and $g \in W^{1,2}(Q)$. This completes the proof under the stated assumptions, with no lower Hessian bound on g. □

C.2 Sharpness of the source curvature factor

Proof of Theorem C.1, sharpness. Fix α, β, ρ and $\kappa = \alpha - \rho$. First let $Q_R(\mathrm{d}z) = e^{-U_R(z)}\,\mathrm{d}z/Z_R$, where

$$U_R(z) = \begin{cases} \beta z^2/2, & |z| \leq R, \\ \alpha z^2/2 + (\beta - \alpha)R|z| - (\beta - \alpha)R^2/2, & |z| > R. \end{cases} \tag{33}$$

This even C^1 potential has weak second derivative in $[\alpha, \beta]$. Since $U_R \geq \alpha z^2/2$ and $U_R(z) \to \beta z^2/2$, dominated convergence gives $\mathbb{E}_{Q_R} Z = 0$ and $\mathrm{Var}_{Q_R}(Z) \to 1/\beta$. Take $g_b(z) = \rho z^2/2 + bz$, with $b \to +\infty$. Writing $c_R = (\beta - \alpha)R$, its positive-tail integral is

$$\int_R^\infty e^{g_b - U_R}\,\mathrm{d}z = e^{c_R R/2 + (b - c_R)^2/(2\kappa)} \int_R^\infty e^{-\kappa[z - (b - c_R)/\kappa]^2/2}\,\mathrm{d}z.$$

For fixed R, the last integral tends to $\sqrt{2\pi/\kappa}$. Conversely, $\int_{\mathbb{R}} e^{g_b - U_R}\,\mathrm{d}z \leq \sqrt{2\pi/\kappa}\, e^{b^2/(2\kappa)}$. Symmetry makes $\mathbb{E}_{Q_R} g_b$ independent of b, so

$$J_{Q_R}(g_b) = \frac{b^2}{2\kappa} + O_R(b), \qquad \mathrm{Var}_{Q_R}(g_b) = b^2\,\mathrm{Var}_{Q_R}(Z) + \frac{\rho^2}{4}\,\mathrm{Var}_{Q_R}(Z^2).$$

Thus

$$\lim_{R\to\infty} \lim_{b\to\infty} \frac{J_{Q_R}(g_b)}{\mathrm{Var}_{Q_R}(g_b)} = \frac{\beta}{2\kappa}. \tag{34}$$

To obtain smooth admissible sources, convolve U_R with a fixed even smooth compactly supported probability density and subtract its value at zero. The result is even, smooth, and has curvature in $[\alpha, \beta]$. Outside bounded transitions around $\pm R$, its quadratic and linear tail coefficients are unchanged; only constants change. For each fixed z, the centered smoothed potential eventually equals $\beta z^2/2$. Evenness and lower curvature give the same dominating Gaussian envelope. The variance limit and the fixed-R leading gap therefore persist, proving (34) within the smooth class. The quadratic log ratios have finite Sobolev norm under these sources. □

D A REMOTE WALL: VARIANCE AND FISHER CERTIFICATES

Let $Q = N(0,1)$, write φ, Φ for its density and distribution function, and set $\overline{\Phi} = 1 - \Phi$.

Proposition D.1 (A separation at positive approximation error)**.** *Fix $b \geq 0$, $0 < \rho < 1$, $\kappa = 1 - \rho$, and define*

$$q_t = \overline{\Phi}(t), \quad a_t = \sqrt{t/q_t}, \quad w_t(z) = (z-t)_+, \quad g_t(z) = bz + \rho z^2/2 - a_t w_t(z), \qquad t > 0.$$

The tilt $P_t = e^{g_t} Q / \mathbb{E}_Q e^{g_t}$ has weak curvature at least κ, with no larger uniform lower bound. Put $\mu_t = \varphi(t) - tq_t$, $\nu_t = (1+t^2)q_t - t\varphi(t)$, and $V_0 = b^2 + \rho^2/2$. Then

$$V_t := \mathrm{Var}_Q(g_t) = V_0 + a_t^2(\nu_t - \mu_t^2) - 2a_t[bq_t + \rho\varphi(t)/2], \tag{35}$$

$$I_t := \mathbb{E}_Q|g_t'|^2 = b^2 + \rho^2 + t - 2a_t[bq_t + \rho\varphi(t)], \tag{36}$$

$$V_t \leq V_0 + 2/t, \qquad V_t \to V_0, \qquad I_t = t + b^2 + \rho^2 + o(1). \tag{37}$$

Moreover,

$$J_Q(g_t) \longrightarrow J_0 := \frac{b^2}{2\kappa} - \frac{1}{2}\log\kappa - \frac{\rho}{2} > 0, \tag{38}$$

$$\mathrm{KL}(Q\|P_t) \leq \frac{V_0 + 2/t}{2\kappa}. \tag{39}$$

Proof. The target potential is $\kappa z^2/2 - bz + a_t(z-t)_+$, up to a constant. Its distributional second derivative is $\kappa\,\mathrm{d}z + a_t\delta_t$, equal to κ off the wall. The log ratio belongs to $W^{1,2}(Q)$. Gaussian integration gives

$$\mathbb{E}_Q w_t = \mu_t, \quad \mathbb{E}_Q w_t^2 = \nu_t, \quad \mathrm{Cov}_Q(Z, w_t) = q_t, \quad \mathrm{Cov}_Q(Z^2, w_t) = \varphi(t).$$

For the covariances, integrate by parts using $w_t' = \mathbf{1}_{\{z>t\}}$ and $w_t'' = \delta_t$. Expansion of the variance proves (35). Expansion of $g_t' = b + \rho z - a_t\mathbf{1}_{\{z>t\}}$, with $\mathbb{E}[Z\mathbf{1}_{\{Z>t\}}] = \varphi(t)$ and $a_t^2 q_t = t$, proves (36).

The tail-shift inequality $\overline{\Phi}(t+u)/q_t \leq e^{-tu-u^2/2} \leq e^{-tu}$, $u \geq 0$, and layer-cake integration give

$$\mu_t \leq q_t/t, \qquad \nu_t \leq 2q_t/t^2, \qquad a_t\mu_t \leq \sqrt{q_t/t} \to 0, \qquad a_t^2\nu_t \leq 2/t \to 0.$$

Hence $a_t w_t \to 0$ in $L^2(Q)$ and $V_t \to V_0$. The subtracted covariance in (35) is nonnegative, giving the upper bound. Mills' asymptotic $q_t \sim \varphi(t)/t$ implies $a_t q_t, a_t\varphi(t) \to 0$ and proves the Fisher limit. Finally, $e^{g_t} \leq e^{bz + \rho z^2/2}$, an integrable envelope, so dominated convergence and $\mathbb{E}_Q g_t = \rho/2 - a_t\mu_t$ give the gap limit. At finite t, completing the square on each side of the wall gives

$$\begin{aligned} M_t := \mathbb{E}_Q e^{g_t} &= \kappa^{-1/2} e^{b^2/(2\kappa)} \Phi(\sqrt{\kappa}\,t - b/\sqrt{\kappa}) \\ &\quad + \kappa^{-1/2} e^{a_t t + (b-a_t)^2/(2\kappa)} \overline{\Phi}(\sqrt{\kappa}\,t - (b-a_t)/\sqrt{\kappa}), \end{aligned} \tag{40}$$

and $J_Q(g_t) = \log M_t - \rho/2 + a_t\mu_t$.

To apply the smooth theorem, replace w_t by $s_\eta(z-t) = \eta\log(1 + e^{(z-t)/\eta})$. For fixed t, convexity preserves the upper Hessian bound, the log ratio is smooth and in $W^{1,2}(Q)$, and its uniform difference from g_t is at most $a_t\eta\log 2$. Theorem A.1 passes to the limit $\eta \downarrow 0$ in the normalizer, mean, and variance, proving the certificate. □

For $b = 1$, $\rho = \kappa = 1/2$, the limiting gap is $3/4 + (\log 2)/2$, the variance certificate tends to $9/8$, and the Fisher certificate is $t + 5/4 + o(1)$. There is also a direct Gaussian-envelope bound: since $g_t \le g_0 := bz + \rho z^2/2$,

$$J_Q(g_t) \le \log \mathbb{E}_Q e^{g_0} - \mathbb{E}_Q g_t = J_0 + a_t \mu_t \longrightarrow J_0. \tag{41}$$

It is asymptotically exact. Thus this family separates the variance and Fisher certificates, while a direct normalizer envelope also succeeds.

Proposition D.2 (Smooth replacement and a rare fourth moment)**.** *Replace w_t by $s_{\eta_t}(z - t)$, with $\eta_t = a_t^{-2}$, and call the log ratio $\widetilde{g}_t$. It is smooth, $\widetilde{g}_t'' \le \rho$, belongs to $W^{1,2}(Q)$, and satisfies*

$$J_Q(\widetilde{g}_t) \to J_0, \qquad \mathrm{Var}_Q(\widetilde{g}_t) \to V_0, \qquad \mathbb{E}_Q|\widetilde{g}_t'|^2/t \to 1.$$

For $\delta_t = (\log 2)/a_t$,

$$\mathrm{Var}_Q(\widetilde{g}_t) \le \left(\sqrt{V_0 + 2/t} + \delta_t\right)^2, \tag{42}$$

$$\mathbb{E}_Q[(a_t w_t(Z))^4] \sim \frac{24}{t^2 q_t} \to \infty. \tag{43}$$

Proof. Convexity of softplus preserves curvature, and $0 \le g_t - \widetilde{g}_t \le \delta_t \to 0$. The gap limit follows from this uniform bound; the centered L^2 triangle inequality gives the variance bound and limit. For $\sigma(u) = (1 + e^{-u})^{-1}$, substitution $z = t + \eta_t u$ and $a_t^2 \eta_t = 1$ give

$$\mathbb{E}_Q|\widetilde{g}_t' - g_t'|^2 \le \varphi(0) \int_{\mathbb{R}} [\sigma(u) - \mathbf{1}_{\{u>0\}}]^2 \, \mathrm{d}u = \varphi(0)(2\log 2 - 1).$$

Thus the two gradient L^2 norms differ by a bounded amount, which proves the Fisher limit. Finally,

$$\mathbb{E}_Q w_t^4 = \varphi(t) t^{-5} \int_0^\infty v^4 e^{-v - v^2/(2t^2)} \, \mathrm{d}v \sim 24 \varphi(t) t^{-5}.$$

Multiplying by $a_t^4 = t^2/q_t^2$ and using Mills' asymptotic proves the fourth-moment assertion. □

The Gaussian envelope also gives $J_Q(\widetilde{g}_t) \le J_0 + a_t \mu_t + \delta_t \to J_0$. The population separation is not an automatic sampling advantage: the fourth moment diverges, so an empirical variance need not be a reliable upper bound. The analytic estimate $a_t^2 \mathbb{E}_Q w_t^2 \le 2/t$ supplies the rare-tail control used here.

E Boundaries of the certificate

Proposition E.1 (Zero curvature margin)**.** *No finite constant bounds $J_Q(g)/\mathrm{Var}_Q(g)$ uniformly over smooth log-concave tilts of $Q = N(0, 1)$, even when each log ratio has a finite global Lipschitz constant.*

Proof. Fix $b > 0$. For $L > 0$, let $g_{L,b}(z) = q_L(z + b)$, where

$$q_L(u) = \begin{cases} 0, & u \le 0, \\ u^2/2, & 0 < u \le L, \\ Lu - L^2/2, & u > L. \end{cases}$$

Its derivative is the clipped ramp $\min\{(z + b)_+, L\}$, so $0 \le g_{L,b}'' \le 1$ weakly. Splitting the Gaussian integral at $-b, L - b$ gives the three terms $\Phi(-b)$, $\varphi(b)(e^{bL} - 1)/b$, and $e^{bL}\Phi(b)$; hence

$$\mathbb{E}_Q e^{g_{L,b}} = e^{bL}[\Phi(b) + \varphi(b)/b] + \Phi(-b) - \varphi(b)/b. \tag{44}$$

Also $0 \le g_{L,b} \le (Z + b)_+^2/2$ and the log ratios converge pointwise to that nonconstant $L^2(Q)$ envelope. Thus $J_Q(g_{L,b}) = bL + O_b(1)$, while the variance converges to a finite positive limit. Smooth convolution of width $\eta_L = L^{-2}$, with a nonnegative mollifier supported on $[-\eta_L, \eta_L]$, preserves the weak curvature bounds and has uniform error at most $L\eta_L \to 0$. The smooth log ratios therefore have the same gap growth and variance limit. □

Proposition E.2 (Curvature above one)**.** *For every $\kappa > 1$, the log ratio $g(z) = -(\kappa-1)z^2/2$ has a target of curvature κ, but*

$$\frac{J_Q(g)}{\mathrm{Var}_Q(g)} = \frac{\kappa - 1 - \log\kappa}{(\kappa-1)^2} > \frac{1}{2\kappa}. \tag{45}$$

Proof. Gaussian integration gives $J_Q(g) = (\kappa - 1 - \log\kappa)/2$ and $\mathrm{Var}_Q(g) = (\kappa-1)^2/2$. The numerator minus $(\kappa-1)^2/(2\kappa)$ is $F(\kappa) = \kappa/2 - \log\kappa - 1/(2\kappa)$, with $F(1) = 0$ and $F'(\kappa) = (\kappa-1)^2/(2\kappa^2) > 0$. □

Such a target can still use the scalar theorem with curvature lower bound one, or the degree-dependent certificate with its full curvature.

F Posterior-certification procedures and controls

F.1 Model, parameterization, and deterministic enclosures

For rows $x_i^\top$ of the whitened training design and binary responses y_i, the target potential is

$$U(\theta) = \tfrac{1}{2}\theta^\top\Lambda\theta + \sum_{i=1}^{n}\{f(x_i^\top\theta) - y_i x_i^\top\theta\}, \qquad f(t) = \log(1+e^t), \quad \Lambda = (n/g_{\mathrm{p}})I.$$

All predictors and the intercept receive a proper prior. Spambase subset seeds are $20, 21, 22$; Covertype seeds are $30, 31, 32$. Each seed uses nested, label-independent permutation prefixes for the two sample sizes. Means, scales, and the whitening transform use only the selected training design. Covertype retains the SVD directions whose singular values exceed the largest singular value times $\max(n, d_{\mathrm{raw}})$ times machine precision. Its posterior is defined on that identified span, not on all dependent raw coefficients. Laplace uses the posterior mode and inverse Hessian; full Gaussian VI minimizes the ordinary reverse-KL objective, initialized at Laplace. VI uses one-dimensional Gaussian quadrature for projected logits, with order 48 for fitting and order 96 for diagnostics. Numerical stopping status is recorded separately from certification.

Let $c_i = L^\top x_i$, $s_i = \|c_i\|$, $a_i = x_i^\top\mu$, and $u_i = c_i/s_i$ when $s_i > 0$. Choose a univariate polynomial p_i approximating $f(a_i + s_i t)$, and replace each softplus factor by $p_i(u_i^\top z)$ to obtain a polynomial P for the log ratio. Zero-scale factors are constant. Hermite coefficients are chosen by order-96 quadrature; their quadrature accuracy is *not* a certificate assumption, since the actual residual of the chosen polynomial is enclosed afterwards. With probabilists' Hermite polynomials H_k, write

$$P = p_0 + \ell^\top z + (z^\top D z - \operatorname{tr} D) + \sum_{k=3}^{m}\sum_{i} b_{ik} H_k(u_i^\top z), \qquad W = A^{-1}.$$

The Gaussian contraction formulas give exact polynomial moments:

$$V_P = \|\ell\|^2 + 2\|D\|_{\mathrm{F}}^2 + \sum_{k=3}^{m} k! \sum_{i,j} b_{ik}b_{jk}(u_i^\top u_j)^k, \tag{46}$$

$$I_P = \ell^\top W\ell + 4\operatorname{tr}(DWD) + \sum_{k=3}^{m} kk! \sum_{i,j} b_{ik}b_{jk}(u_i^\top W u_j)(u_i^\top u_j)^{k-1}. \tag{47}$$

These contractions are evaluated in row blocks, without constructing high-order tensors. Taylor degrees $3, 5$ and Hermite degrees $5, 8$ are supplied symmetrically to both certificates. The original Spambase Taylor variants use global derivative remainders; the subsequent Hermite variants are tighter for both methods on all 24 fits. Covertype supplies the actual-residual refinement below to all four variants. Thus the comparison does not attribute an unequal numerical refinement to the population inequality.

Define residuals $R_i(t) = f(a_i + s_i t) - p_i(t)$ and $r_i(t) = f'(a_i + s_i t) - p_i'(t)/s_i$. For upper bounds $v_i \geq \mathbb{E}R_i(Z)^2$, $w_i \geq \mathbb{E}r_i(Z)^2$, $Z \sim N(0,1)$, put

$$r_V = \sum_i \sqrt{v_i}, \qquad r_F = \sum_i \sqrt{c_i^\top W c_i}\sqrt{w_i}.$$

Minkowski and its reverse, together with Theorem 3.1 and the target log-Sobolev inequality, yield

$$B_V = \frac{(\sqrt{V_P} + r_V)^2}{2\kappa}, \qquad B_F = \frac{(\sqrt{I_P} + r_F)^2}{2}, \qquad \underline{B}_F = \frac{(\sqrt{I_P} - r_F)_+^2}{2}. \tag{48}$$

Here $J \le B_V, B_F$, while $\underline{B}_F \le \frac{1}{2}\mathbb{E}\nabla g^\top W \nabla g$. The last quantity is a lower bound on a Fisher certificate, never on KL. Correlations between projected logits do not affect these inequalities. Minima of valid deterministic upper bounds and maxima of lower bounds over the retained variants remain valid.

The residual integrals use 4096 equal intervals on $[-10, 10]$. For either residual v, on a cell of width h and midpoint t_0, let $M_2 \ge \sup|v''|$. Set

$$B_1 = |v'(t_0)| + hM_2/2, \quad B_0 = |v(t_0)| + h|v'(t_0)|/2 + h^2 M_2/8.$$

Then $\sup|(v^2)''| \le 2(B_1^2 + B_0 M_2)$. Integrating the endpoint linear interpolant of v^2 against the Gaussian density and adding $h^2(B_1^2 + B_0 M_2)/4$ times the cell probability encloses that cell's integral. Gaussian hat integrals use only Φ and φ. Bounds $f'' \le 1/4$, $|f'''| \le 1/(6\sqrt{3})$, and absolute polynomial derivative coefficients provide M_2. For arbitrary chosen polynomials, the exterior uses $f(a + st) \le \log 2 + |a| + s|t|$, $0 < f' < 1$, and the polynomial coefficients, with the Gaussian tail-moment recurrence

$$T_0(R) = 2\overline{\Phi}(R), \qquad T_{2k}(R) = 2R^{2k-1}\varphi(R) + (2k-1)T_{2k-2}(R).$$

Taking the squared sum of coefficient-weighted $\sqrt{T_{2k}(10)}$ bounds each exterior residual square integral. This argument requires neither exact Hermite projection nor independent factors. Ordinary floating evaluation, including cancellation in residuals and Gram contractions, remains distinct from an outward-rounded machine proof.

F.2 Posterior-mass conversion

For $Q = N(\mu, LL^\top)$, define $E_q = \{\theta : \left\|L^{-1}(\theta - \mu)\right\| \le F_{\chi_d}^{-1}(q)\}$. Binary-KL data processing gives

$$\mathrm{kl}(q\|P(E_q)) \le \mathrm{KL}(Q\|P) \le B, \qquad P(E_q) \ge b_q(B),$$

where $b_q(B)$ is the lower root of $\mathrm{kl}(q\|p) = B$. For $q = 0.995$, the 0.95 target permits $B \le 0.0345364233362$. Historical region levels $q \in \{0.95, 0.99, 0.995, 0.999\}$ and desired masses $0.9, 0.95, 0.99$ were chosen after the KL study and before this conversion. The prospective study retains the displayed pair.

Both methods also receive a transport conversion. The target's anisotropic transport inequality gives $W_{2,A}^2 \le 2B$. For $r = F_{\chi_d}^{-1}(q)$, define $C_r(s) = \int_s^r (r-u)^2\, dF_{\chi_d}(u)$. Solving $C_r(s) = 2B/\lambda_{\min}(A)$ gives the mass lower bound $F_{\chi_d}(s)$, or zero if the budget exceeds $C_r(0)$. We take the better of this conversion and binary KL. The integrals use incomplete-gamma moments. Evaluating the decreasing conversion at a Fisher-functional lower endpoint gives an optimistic Fisher ceiling, not a posterior-mass guarantee. The historical full-metric check uses the radial shell coupling, with cost $(\mathrm{tr}\, A/d)C_r(s)$. That is the cost of one particular coupling, not a lower bound on the optimal transport cost, so it is a diagnostic rather than a guarantee; on the recorded inputs it does not improve the binary-KL conclusion, and no reported decision uses it. These comparisons concern the stated information and conversions, rather than every possible method for estimating event probabilities.

F.3 Prospective within-family study

The row selection, model grid, numerical variants, resource caps and reporting rules were frozen before new fits. The historical usage union contains 12,198 Covertype rows. From the remaining 568,814 rows, a PCG64 permutation with seed 130 selects three disjoint blocks of 4,096 rows; each smaller design uses its first 1,024 rows. The smaller/larger design ranks are 48/51, 48/52 and 46/50. Saved indices verify exclusion and disjointness. The dataset and model family were chosen in earlier work; methods, priors and nested sizes within a block remain paired.

The grid contains 18 targets and 36 planned approximations. At each size, use $g_{\mathrm{p}} = 1, 4$, Laplace and full-Gaussian VI. At the larger size add $g_{\mathrm{p}} = 16$ and prior covariance I in the identified whitened

Table 3: All prospective groups. Entries are the minimum and maximum computed upper bounds across the three blocks, using each method's best of the four fixed variants. Missing fits stay in the planned denominator. These ranges describe fixed certificate values, not confidence intervals for a population success rate.

n	Prior	Approximation	Computed	Variance	Matrix Fisher
1024	$g_{\mathrm{p}} = 1$	Laplace	3/3	.002271–.002885	.004604–.005897
1024	$g_{\mathrm{p}} = 1$	VI	3/3	.000603–.000756	.002045–.002588
1024	$g_{\mathrm{p}} = 4$	Laplace	3/3	.12941–.15569	.21356–.25987
1024	$g_{\mathrm{p}} = 4$	VI	3/3	.041944–.051448	.12452–.15477
4096	$g_{\mathrm{p}} = 1$	Laplace	3/3	.001258–.001609	.002493–.003286
4096	$g_{\mathrm{p}} = 1$	VI	3/3	.000362–.000439	.001186–.001478
4096	$g_{\mathrm{p}} = 4$	Laplace	3/3	.076050–.092050	.12707–.15338
4096	$g_{\mathrm{p}} = 4$	VI	3/3	.025055–.030973	.073475–.090432
4096	$g_{\mathrm{p}} = 16$	Laplace	3/3	1.9669–2.4409	2.1667–2.6994
4096	$g_{\mathrm{p}} = 16$	VI	3/3	.64456–.81047	1.3170–1.6533
4096	$\Sigma_{\mathrm{prior}} = I$	Laplace	3/3	$[1.94, 2.45] \times 10^6$	48,830–65,931
4096	$\Sigma_{\mathrm{prior}} = I$	VI	0/3	–	–

coordinates. All four polynomial variants are assigned to each fit, with the residual settings given above. The budgets are 300 process CPU seconds per fit and 120 per variant; cases are not replaced after nonattainment.

There are 33 returned fits and 132 computed certificate pairs out of 144 planned. All three unit-prior-variance VI fits reach the fitting cap. The twelve $g_{\mathrm{p}} = 1, 4$ VI fits pass the order-96 fitting diagnostic; the three $g_{\mathrm{p}} = 16$ VI fits fail it despite solver convergence. Their fixed-reference bounds and diagnostic status remain in the record. Original KL thresholds are $0.001, 0.01, 0.05, 0.1$.

For weak priors, the reverse triangle inequality also gives

$$\frac{(\sqrt{\mathrm{Var}_Q(h)} - R)_+^2}{2\kappa} \leq \frac{\mathrm{Var}_Q(g)}{2\kappa}.$$

On the saved $g_{\mathrm{p}} = 16$ VI cases, the corresponding lower endpoints are $0.57245, 0.56524, 0.46013$; Laplace endpoints range from 1.63809 to 2.01595. Subject to the same floating qualification, these scalar functionals cannot reach 0.1 through more accurate variance evaluation alone. This does not lower-bound KL or exclude another reference or certificate.

F.4 Degree-dependent reanalysis

The refined algorithm was developed after the original outcomes and reuses all 84 planned cases. Its 336 planned variant slots contain 276 computed variants: 48 refined Taylor variants were not saved for historical Spambase, and twelve belong to the censored fits. No missing variant is replaced. Each method takes its best eligible upper at the original thresholds. The reference pool combines scalar variance, matrix Fisher and the eligible basic OU certificate. The basic Ornstein–Uhlenbeck (OU) comparator implements $J_Q(g) \leq \mathcal{D}_A(g)/2$, when $0 \prec A \preceq I$, with the same polynomial and residual treatment; $\mathcal{D}_A$ is defined in Appendix B.

The positive scalar integrals in the degree weights are evaluated at 40 and 70 digits. Very small components retain a conservative Fisher contribution; measured low-degree terms are not set to zero. The refinement adds four Spambase KL decisions and six Covertype region decisions to that pool. All ten gains select degree-eight Hermite polynomials. A separately specified audit reconstructs geometry and low-degree coefficients at 80 digits, higher-degree contractions and all factor residuals in wider arithmetic, with corrections for differing polynomial representations. All ten decisions survive; the largest change in an upper bound is below 2.4×10^{-14}. A later check of inconsistent saved maximum-eigenvalue fields leaves every basic-OU eligibility decision unchanged. These checks support numerical stability, not outward rounding or fresh confirmation of the algorithm. In the reported exploratory values, the stronger comparison below removes six region and two KL increments; the unresolved common-target/reference qualification in Appendix F.5 prevents treating this as a verified same-target comparison.

F.5 THE STRONGER GAUSSIAN PRIOR-ALLOCATION CONTROL

The following classical geometric-integral construction (Bennett et al., 2005) is the stronger control of Section 6.4. It allocates prior curvature to the likelihood factors, applies the geometric Brascamp–Lieb inequality, and optimizes a uniform allocation scalar with fixed moment-based weights. Both families receive the same Laplace and VI donor pool for each target. It was evaluated only on 24 existing $n = 1024$ references, drawn from both the historical and the prospective designs (the panel is specified below).

Let $f(s) = \log(1 + e^s)$, and write the unnormalized potential as

$$U(\theta) = \tfrac{1}{2}\theta^\top \Lambda\theta + \sum_i [f(x_i^\top\theta) - y_i x_i^\top\theta], \qquad \Lambda = pI, \quad p > 0.$$

Assume X has full column rank and no zero rows. Fix a donor Gaussian $Q = N(\mu, LL^\top)$ and positive weights α_i and finite β_i. The implementation uses fixed Gaussian quadrature approximations to $\mathbb{E}_Q f''(x_i^\top\theta)$ and $\mathbb{E}_Q f'(x_i^\top\theta)$. Their approximation error does not affect the algebraic validity of using them as arbitrary fixed weights. Set

$$\begin{aligned} &G = X^\top X, && M = \Lambda + X^\top \operatorname{diag}(\alpha) X, && BB^\top = M^{-1}, \\ &r = \Lambda\mu + X^\top(\beta - y), && t = \beta - XG^{-1}r, && \eta = X\mu, \end{aligned}$$

where B is positive-diagonal Cholesky. Thus $X^\top t = X^\top y - \Lambda\mu$ without assuming VI stationarity. For $0 < \rho \le \rho_* = p/\lambda_{\max}(G)$, let

$$h_i = \|B^\top x_i\|^2, \quad c_i = (\rho + \alpha_i)h_i, \quad u_i = B^\top x_i/\sqrt{h_i}, \quad K = B^\top(\Lambda - \rho G)B \succeq 0.$$

The frame identity is $\sum_i c_i u_i u_i^\top + K = I$. Define

$$I_i(\rho) = \int_{\mathbb{R}} \exp\left[\frac{-\rho h_i z^2/2 + t_i\sqrt{h_i}z - f(\eta_i + \sqrt{h_i}z)}{c_i}\right] \mathrm{d}z.$$

These integrals are finite because $\rho > 0$. With $C_0 = -\mu^\top\Lambda\mu/2 + y^\top\eta$, the geometric Brascamp–Lieb inequality gives

$$\log Z \le \mathcal{U}(\rho) := C_0 + \log\det B + \frac{\operatorname{tr} K}{2}\log(2\pi) + \sum_i c_i \log I_i(\rho). \tag{49}$$

Indeed, substitute $\theta = \mu + Bz$. The likelihood factors above and the one-dimensional Gaussian factors associated with the positive eigenvalues of K reconstruct the integrand exactly. The frame identity then bounds its integral by the product of scalar integrals. Zero eigenweights at ρ_* are omitted. Adding the recipient's complete free energy $\mathbb{E}_Q U - H(Q)$ converts (49) into a KL upper. Each donor's normalizer upper can be transferred to another reference for the identical target, including the prior; its KL upper cannot be transferred unchanged.

Optimizing the uniform scalar allocation. Keep the donor, α, and β fixed as ρ changes. Put

$$A_i(z) = \alpha_i h_i z^2/2 + t_i\sqrt{h_i}z - f(\eta_i + \sqrt{h_i}z).$$

Since $\sum_i c_i + \operatorname{tr} K = d$, the variable part of $\mathcal{U}(\rho)$ is $\sum_i c_i \log \mathbb{E}_{\gamma_1}\exp(A_i/c_i)$. For $F(c) = c\log\mathbb{E}_{\gamma_1} e^{A/c}$ and the corresponding tilted measure ν_c, direct differentiation yields

$$F'(c) = -\operatorname{KL}(\nu_c\|\gamma_1) \le 0, \qquad F''(c) = \operatorname{Var}_{\nu_c}(A)/c^3 \ge 0.$$

Positive ρ provides Gaussian tails that justify these derivatives. Thus $\mathcal{U}$ is decreasing and convex; continuity and the semidefinite frame identity give its minimum at ρ_*. This optimizes one scalar with frozen moment-based weights, not arbitrary row allocations, weights, donors, or grouped factors.

For $0 < \rho_0 \le \rho_*$, an explicit error bound permits evaluation strictly inside the domain. The negative log-density of the ith scalar factor has second derivative between $m_i = \rho_0 h_i/c_i$ and $M_i = (\rho_0 + 1/4)h_i/c_i$. Its mode satisfies $|z_i| \le \max\{|t_i - 1|, |t_i|\}/(\rho_0\sqrt{h_i})$. Integration by

parts gives $\mathbb{E}(Z - z_i)^2 \le 1/m_i$, while the upper curvature bound gives $H(\nu_i) \ge \frac{1}{2}\log(2\pi/M_i)$. Consequently $-\mathcal{U}'(\rho_0) \le D(\rho_0)$, where

$$D(\rho_0) = \sum_i \left[\frac{\max\{|t_i - 1|, |t_i|\}^2}{\rho_0^2} + \frac{c_i}{\rho_0} + \frac{h_i}{2} \log \frac{(\rho_0 + 1/4)h_i}{c_i} \right].$$

Convexity and monotonicity imply

$$\mathcal{U}(\rho_0) - (\rho_* - \rho_0)D(\rho_0) \le \mathcal{U}(\rho_*) \le \mathcal{U}(\rho_0).$$

The left endpoint bounds the optimized comparator functional, not the true log normalizer or KL divergence. Taking minima of both endpoints over the complete frozen donor pool preserves this distinction.

The frozen panel comprises both Laplace and VI references for six Covertype blocks at $n = 1024, g_{\mathrm{p}} = 4$ (the three historical subsets and the three prospective blocks), and three Spambase subsets at $n = 1024, g_{\mathrm{p}} \in \{1, 4\}$. All 24 references are reused; there is no refitting. We evaluate $\rho_0 = (1 - 10^{-10})\rho_*$ with fixed order-96 moment weights, 96/192-node integral and free-energy diagnostics, and an independent adaptive quadrature implementation for every scalar factor. All 24 donors meet weighted discrepancy/error-estimate gates of 10^{-8}, and all optimized-functional brackets have width at most 10^{-6}. These are floating diagnostics, not directed enclosures.

Every donor normalizer upper is combined with the recipient's complete free energy before taking the best bound. Both methods select the VI donor on all twelve target–prior pairs. The historical degree computation used $\mathrm{chol}(\Sigma)$; the later allocation computation used $\mathrm{chol}((\Sigma + \Sigma^\top)/2)$. Matching saved means, covariances, and transformation recipes does not establish exact common-target/reference identity across these computations. A valid comparison or transfer requires the same target, including the prior, and an appropriately controlled recipient free energy or free-energy difference. That complete verification has not been established. We therefore treat the entire 24-reference panel as exploratory, including both datasets, both reported method orderings, and all apparent threshold crossings.

Table 4: **Exploratory stronger-control comparison on Spambase, $n = 1024$, VI.** The table contains all four KL decisions added by the degree-resolved certificate to the original variance/Fisher/basic-OU pool. "Best control" is the minimum of that pool and the scalar-optimized classical bound. Bold entries identify the two apparent threshold crossings missed by the best control in the rounded floating values. Exact common-target/reference assurance and a qualified free-energy comparison remain unresolved for the entire panel; neither these crossings nor either method ordering is counted as a verified same-target result.

Seed	g_{p}	Target	Degree	Classical	Best control
20	4	0.05	0.039836	0.047405	0.047405
21	1	0.001	**0.000911**	0.001233	0.001012
21	4	0.05	0.043262	0.048425	0.048425
22	4	0.05	**0.049444**	0.052941	0.052941

In the reported exploratory values, the classical bound is smaller on all twelve Covertype references and removes the six additional region decisions of the degree-resolved certificate. On Spambase, the shared degree bounds are smaller, and two apparent crossings in Table 4 satisfy the original threshold rule. At those crossings, the degree bounds lie 8.9% and 1.1% below their targets, and the best controls lie 1.2% and 5.9% above. Both methods select the VI donor for each target–prior pair; transferring that donor's normalizer bound to its paired Laplace reference does not create an independent success. Because the common-target/reference and free-energy qualification applies to the whole panel, none of these observations is promoted to a verified same-target comparative result. Establishing such a result requires re-evaluation from a canonical record or a qualified bridge between the representations.

The retrospective stricter diagnostic requires degree ≤ 0.95 times its target, comparator ≥ 1.05 times the target, and degree ≤ 0.90 times the best comparator. None of the 24 references satisfies all three. The combined pool includes the original bounds and scalar-optimized classical control, but not a complete newer direct/stochastic suite. A lower endpoint for an optimized comparator functional is not a lower bound on true KL.

F.6 NUMERICAL VALIDATION

Eight predeclared logistic targets in one and two dimensions test the normalizer, variance and matrix-Fisher calculations independently, using physical-coordinate Gauss–Legendre integration and Gaussian integration in covariance-eigenvector coordinates. All four polynomial/residual variants are checked. Nine Gaussian controls have closed-form answers; selected actual residual integrals are evaluated at 40 and 70 digits. The last prescribed normalizer orders differ by at most 4.19×10^{-11}. These are convergence diagnostics, not complete enclosures of rounding and integration error.

The two historical passing KL decisions with the smallest relative margins were separately selected before wider-arithmetic recomputation: Covertype $n = 1024, g_{\mathrm{p}} = 4$, VI, seeds 32 and 30. Their degree-eight variance uppers, 0.049190484446978884 and 0.048961863967965834, remain below 0.05; each changes by less than 2.7×10^{-14}, and its Fisher lower endpoint remains above 0.05. This check covers the saved fixed Gaussians and polynomials, rather than exact high-dimensional KL or every experiment.

F.7 ADDITIONAL COMPARISONS AND RETAINED EXPLORATORY OUTCOMES

We retain comparisons that delimit the method's practical use. Their complete protocols, sources, cases and failed runs remain in the numerical artifact; they introduce no additional positive claim.

Poisson factor control and cost. For the working potential $U(\theta) = \theta^\top \Lambda\theta/2 + \sum_i e_i e^{x_i^\top \theta} - \sum_i y_i x_i^\top \theta$, $e_i > 0$, and $Q = N(m, S)$, set

$$w_i = e_i e^{x_i^\top m + x_i^\top S x_i/2}, \quad M = \Lambda + \sum_i w_i x_i x_i^\top, \quad r = \sum_i (w_i - y_i) x_i, \quad h_i = x_i^\top M^{-1} x_i.$$

Geometric Brascamp–Lieb with likelihood and compensating Gaussian-prior factors, followed by the Gamma-integral identity and $0 < R(t) < 1/(12t)$ for the log-Gamma remainder (National Institute of Standards and Technology, 2026), gives

$$\mathrm{KL}(Q\|P) \le \tfrac{1}{2}\{\mathrm{tr}(MS) - d - \log\det M - \log\det S\} + \tfrac{1}{2}(\Lambda m + r)^\top \Lambda^{-1}(\Lambda m + r) + \tfrac{1}{12}\sum_i w_i h_i^2. \tag{50}$$

For completeness, take $c_i = w_i h_i$ and $u_i = M^{-1/2} x_i / \sqrt{h_i}$. The residual frame is $M^{-1/2}\Lambda M^{-1/2}$; its eigenvectors give Gaussian factors. The likelihood potential is $e_i e^{\sqrt{h_i}s}/c_i - w_i\sqrt{h_i}s/c_i$, whose integral is $\Gamma(1/h_i)(c_i/e_i)^{1/h_i}/\sqrt{h_i}$. The compensating prior linear term contributes $r^\top \Lambda^{-1} r/2$ to the log-normalizer bound. Expanding log-Gamma, subtracting the Gaussian ELBO and bounding its remainder gives (50). Zero design rows are split as constants. Neither nonnegative fitting residual is dropped.

The Poisson targets use two count datasets distributed with the R package MASS (Venables & Ripley, 2002). The quine data (Aitkin, 1978) record days absent from school for 146 children with four categorical covariates; the ships data (McCullagh & Nelder, 1983) record damage incidents by ship type, construction period and operation period, with aggregate months of service as exposure. Each model is a Poisson log-linear regression with categorical main effects and a Gaussian prior, at two prior scales. On the resulting four target–prior pairs, each with Laplace and VI references, this factor bound is tighter than the tested degree and stochastic bounds. All four Quine references and both Ships VI references pass 0.05; Ships Laplace remains uncertified. The primary computations use directed 80-digit arithmetic on exact represented inputs; the factor calculation has a separate 100-digit diagnostic.

The subsequent cost protocol runs fresh single-thread processes serially in fixed alternating order, three times per pair. Each bank evaluates both donors and all transfers; degree uses cutoffs $4, 8, 12$. It includes startup, reconstruction, method work, internal checks, persistence and teardown. Factor has no-worse uppers and lower reported production CPU in every repeat. Quine/prior-sd-4 median CPU is 9.439421 seconds for degree and 2.538625 for factor. The runtime uses Python 3.12.3, NumPy 2.4.6, C Decimal and mpmath 1.3.0 on Linux; a subsequent same-host inspection identifies an AMD EPYC 7R13 CPU. A fixed 0.02-second comparison guard is applied to each total. Fitting, infrastructure and full observer lifetime are excluded or unmeasured. Repetitions reuse the same inputs; they establish the reported implementation comparison, not an end-to-end or general speed ranking.

Complete radial diagnostic. All twelve saved Spambase references in the newer panel are tested against a radial sufficient condition for strict separation of the variance and degree certificates from an optimized family of Gaussian factor bounds; six VI references are primary. The complete own-reference degree and variance pools both yield zero strict and zero material witnesses. The frozen material rule requires the native upper to be at most four-fifths of the witness lower, with an absolute difference at least 10^{-6}, spanning both priors and at least two blocks. All references and numerical controls complete, with an independent Gaussian-moment/100-digit scalar check. This is failure of that sufficient witness, not evidence that the optimized comparator is actually tighter. The fitted references are reused and their inputs retain the original qualified floating assurance.

Selection history. The logistic grids were preceded by two exploratory studies. On a three-parameter Banknote model with 36 approximations, variance and matrix Fisher tied at the primary tolerance 0.1 (9/36 each). On WDBC ($d = 31$, prior scales $g_{\mathrm{p}} \in \{1, 4, 16, 64\}$), useful certificates arose only at $g_{\mathrm{p}} = 1$. On the selected $n = 512$, $g_{\mathrm{p}} = 1$ cases an apparent threshold advantage disappeared once both methods received degree-eight Hermite polynomials with certified residuals; variance remained about 3.3 times tighter for VI and 1.7 times for Laplace. The prior grid was then narrowed to $g_{\mathrm{p}} \in \{1, 4\}$ for Spambase. Covertype was added afterwards as a separately frozen transfer, followed by the prospective study. Earlier Fourier and predictive constructions whose apparent gains were overturned by stronger controls are retained in the artifact.

A stochastic certifier and cost. On six VI fits with $g_{\mathrm{p}} = 1$ (Spambase $n = 4096$, Covertype $n = 1024$, three subsets each), a direct Monte Carlo certifier was run to the same tolerance 0.001. It uses a mixture importance proposal and empirical-Bernstein and bounded-variable KL confidence bounds, with joint 95% confidence. It first certified at $4{,}194{,}304$ proposal draws in every case. Its single-thread CPU time to that checkpoint was 119–187 times the median time of the deterministic path, which itself took about 1–4 seconds per fit including the paired Fisher computation. The cohort was chosen after favorable deterministic results, and the ratio measures the first successful registered checkpoint, not an optimal stopping time. At this tolerance the Fisher functional's lower endpoint exceeds 0.001 in all six cases.

Incomplete and peripheral studies. The broader Spambase comparison has twelve references and 312 required method slots; three endpoint audits per reference remain unresolved, so its primary comparative outcome is unset. The Wine multiclass attempt fails a fitting-convergence gate; a subsequent saved-output diagnostic has four of six endpoint checks converged and no comparative verdict. Neither is counted as a positive or a completed negative efficacy study. A separate eight-case inverse-model screen adds no decisions at its frozen $19{:}1$ posterior-odds target. Earlier Student, perturbed-reference and direct-mixture comparisons are retained in the artifact with their original assumptions, cohort selection and nonattainment. They do not support a stronger claim than the main text.