# Pointwise Complexity for Gaussian Fields: Upper Envelopes, Algorithmic Lower Bounds, and Separation

Yunbei Xu
National University of Singapore
yunbei@nus.edu.sg

**Abstract**

We prove a variance-aware pointwise majorizing-measure theorem for centered Gaussian processes. Classical generic chaining characterizes the scalar quantity $\mathbb{E}\sup_{x\in T} X_x$; the theorem here gives a simultaneous high-probability envelope for the entire field. For an ambient prior $\mu$, the envelope at $x$ is governed by a *pointwise* Fernique–Talagrand functional

$$\Phi_\mu(x) := \int_0^{4\sigma(x)} \sqrt{\log \frac{1}{\mu(B_d(x,\varepsilon))}}\, d\varepsilon,$$

together with the corresponding Gaussian tail term. The theorem is intended as a standard reference whenever one needs a field-level refinement of classical generic chaining, and as the Gaussian-process counterpart of pointwise empirical-process bounds for deep neural networks (Li and Xu, 2026).

We also derive a Bayesian algorithmic lower envelope from the interactive Fano/data-processing principle (Chen et al., 2024). For a known prior $\pi$, an observation channel, and a concrete estimator $\widehat{t}(Y)$, the lower bound is expressed through the exact ghost small-ball mass $\mathbb{E}_{Y\sim Q}\pi(B_d(\widehat{t}(Y),\Delta))$, rather than a worst-case covering number. In Gaussian location experiments, comparison decoders convert Bayes location error into lower bounds on decision-aligned Gaussian ranges. We then construct an elementary weighted-basis example separating the usual Fano relaxation for a fixed prior, the Bayesian algorithmic lower envelope, the pointwise Gaussian envelope on the estimator-selected subatlas, and the full-class minimax risk/global Gaussian scale. Together, these results show that Bayesian algorithmic lower bounds provide local-geometric certificates of pointwise complexity for fixed estimators in overparameterized ambient classes, precisely in regimes where classical minimax theory becomes either too coarse or oracle-dependent.

The appendices apply the framework to finite-cutoff renormalization and graph local time, formalizing analogies between renormalization group flows and deep neural networks, and between graph local time and pointwise dimension. They provide concrete examples where pointwise complexity yields sharper, trajectory-aligned descriptions than global complexity scales.

## Contents

# 1 Introduction and Main Results

Classical generic chaining controls the scalar quantity $\mathbb{E}\sup_{x\in T} X_x$ for a Gaussian process $\{X_x\}_{x\in T}$. For many field-level questions, however, this scalar is not the first object one wants to estimate. In the graph setting, the second Ray–Knight theorem identifies a local-time field at an inverse-local-time stopping rule with a shifted square of a Gaussian free field. In constructive and lattice field theory, one similarly studies local observables before passing to global maxima or continuum limits. This suggests that the Gaussian input should itself be pointwise: one should first prove a simultaneous field-level envelope and only afterwards take suprema or threshold events.

A further motivation comes from recent work on pointwise generalization for deep neural networks (Li and Xu, 2026). There the central object is not a class-wide supremum but a hypothesis-dependent finite-scale complexity. The proof is structural: an exact non-perturbative telescoping expansion exposes learned feature Gram matrices, and a hierarchical local-chart/global-atlas prior converts the resulting data-dependent active subspaces into a data-independent bound. The present paper develops the Gaussian-process counterpart of that viewpoint. Theorem 1.1 can serve as a standard reference whenever one needs a field-level refinement of classical generic chaining.

We also record a single-radius Bayesian lower-envelope counterpart. It is not a pathwise lower bound on each Gaussian coordinate. Rather, it is a data-processing Bayes-risk lower bound for a specified prior, observation channel, and estimator. In Gaussian location experiments, comparison decoders turn Bayes location error into an algorithmic lower bound on a decision-aligned Gaussian range. Conversely, a simultaneous pointwise Gaussian envelope yields an algorithmic upper bound on the same decoding risk. Ordinary nearest-neighbor decoding is the simplest instance; the separating example in Theorem 1.7 uses an unrestricted norm-regularized nearest-neighbor rule over the full ambient class, reflecting the inductive-bias/interpolation principle that good empirical fit must be paired with a locality or complexity bias (Cover and Hart, 1967; Belkin et al., 2018). We then construct an elementary weighted-basis "hub–leaves–cloud" example in which the fixed-prior Fano relaxation is vacuous, the algorithmic ghost-mass lower envelope is sharp for the chosen Bayes problem and for the pointwise upper envelope on the estimator-selected subatlas, while the full-class minimax risk and the global Gaussian supremum scale are much larger.

These results clarify why a Bayesian algorithmic lower bound can provide the right certificate for a pointwise upper envelope under specified priors, even when the ambient class is overparameterized, where classical minimax theory may become either divergent or oracle-dependent. In this sense, the pointwise and algorithmic viewpoint developed here points to the need for decision-theoretic criteria that are localized to the algorithm or estimator and the reference law. This parallels probabilistic refinements of worst-case analysis in theoretical computer science (Yao, 1977; Levin, 1986; Spielman and Teng, 2004). Just as distributional or smoothed input models can make the effective computational complexity of an algorithm much smaller than its worst-case complexity, a concrete estimator and reference law can make the effective probabilistic or information-theoretic complexity of a rich function class much smaller than what is suggested by a global supremum or class-wide relaxation.[1]

We further extend the pointwise-complexity principle to statistical-physics settings, with the details deferred to the appendices. For renormalization maps and interacting fields, we formalize a structural analogy between exponential-family renormalization group (RG) flows and nonlinear feature learning in deep neural networks (DNN). This leads to finite-cutoff pointwise theorems and to a nonconvex phase-atlas example. For Gaussian free fields and continuous-time random walks on finite graphs, we connect local-time and cover-time questions with pointwise-dimension and covering-number scales. The phase-star and decorated phase-star examples show that local time on a target or trajectory-aligned subatlas can have a strictly smaller intrinsic scale than any argument that first passes through full cover time. These appendices are intended as initial results and guideposts for a connection that has not yet been fully explored.

In summary, the paper connects field-level generic chaining with decision-aligned certificates and demonstrates strong separations between pointwise and global complexity scales.

### 1.1 Main pointwise Gaussian theorem

Let $\{X_x\}_{x\in T}$ be a centered Gaussian process on a finite index set $T$. Fix an anchor $x_0 \in T$ such that $X_{x_0} = 0$ almost surely, and define

$$d(x,y) := \left(\mathbb{E}|X_x - X_y|^2\right)^{1/2}, \qquad \sigma(x) := (\mathbb{E}X_x^2)^{1/2} = d(x,x_0), \qquad R_* := \sup_{x\in T} \sigma(x).$$

For an ambient prior $\mu \in \Delta(T)$, where $\Delta(T)$ denotes the set of probability measures on $T$, define the localized pointwise Fernique–Talagrand functional

$$\Phi_\mu(x) := \int_0^{4\sigma(x)} \sqrt{\log \frac{1}{\mu(B_d(x,\varepsilon))}}\, d\varepsilon, \qquad B_d(x,\varepsilon) := \{y \in T : d(x,y) \le \varepsilon\}. \tag{1.1}$$

Here the upper limit is local: it is proportional to the standard deviation $\sigma(x)$, rather than the global radius. The factor 4 is inessential; it absorbs the two constant-factor losses coming from anchoring and from the nearest-point pushforward used in the ambient-equivalence step. If one instead works with the non-local full-radius integral, then the upper limits $R_*$ and $2R_*$ are equivalent up to universal constants, since the entropy integrand is nonincreasing in $\varepsilon$.

Write

$$\Phi_{\mu,*} := \sup_{x\in T} \Phi_\mu(x), \qquad \log_2^+(u) := \max\{\log_2 u, 0\}.$$

The degenerate case $R_* = 0$ is trivial, and the statements below are understood to be vacuous if $\Phi_{\mu,*} = +\infty$.

[1] Here "algorithmic" refers to bounds that depend on a concrete estimator and reference law. This usage is distinct from classical algorithmic information theory, which is traditionally formulated for finite strings or countable description spaces. For related notions of pointwise dimension and their connections to algorithmic and effective statistical dimension, see Li and Xu (2026); Lutz (2016).

For $r_0 > 0$, $s_0 \in (0, R_*]$, and $\delta \in (0,1)$, set

$$M_{r_0,s_0} := \big(1 + \lceil \log_2^+(\Phi_{\mu,*}/r_0)\rceil\big)\big(1 + \lceil \log_2(R_*/s_0)\rceil\big), \tag{1.2}$$

and

$$\mathcal{E}_{\mu,r_0,s_0}(x;\delta) := \max\{\Phi_\mu(x), r_0\} + \max\{\sigma(x), s_0\}\sqrt{\log\left(\frac{eM_{r_0,s_0}}{\delta}\right)}. \tag{1.3}$$

Thus all logarithmic losses from the two successive pointwise-peeling steps are isolated in $M_{r_0,s_0}$.

We state the main result as a variance-aware pointwise Gaussian majorizing-measure theorem in high probability.

**Theorem 1.1** (Pointwise Gaussian majorizing-measure bound)**.** *Assume $0 < R_* < \infty$ and $\Phi_{\mu,*} < \infty$. There exists an absolute constant $C > 0$ such that, for every ambient prior $\mu \in \Delta(T)$, every $r_0 > 0$, every $s_0 \in (0, R_*]$, and every $\delta \in (0,1)$, with probability at least $1-\delta$,*

$$\forall x \in T: \qquad X_x \le C\,\mathcal{E}_{\mu,r_0,s_0}(x;\delta). \tag{1.4}$$

*Consequently, applying the same bound to $X$ and $-X$ and taking a union bound, with probability at least $1-\delta$,*

$$\forall x \in T: \qquad |X_x| \le A_{\mu,r_0,s_0}(x;\delta), \qquad A_{\mu,r_0,s_0}(x;\delta) := C\,\mathcal{E}_{\mu,r_0,s_0}(x;\delta/2). \tag{1.5}$$

*When $r_0, s_0$ are fixed, we abbreviate $A_{\mu,r_0,s_0}$ to $A_\mu$.*

Besides providing the Gaussian-process counterpart of the pointwise empirical-process bound in Li and Xu (2026, Theorem 1), the present result makes a key technical refinement: the global boundedness parameter for the metric is replaced by the local variance scale $4\sigma(x)$. Moreover, in the Gaussian-process setting we do not need the advanced symmetrization machinery or the mixed empirical-ghost metric used in Li and Xu (2026, Theorem 1).

Before presenting the in-expectation consequence, we record the corresponding anchored form of the classical generic-chaining scale. This is a localized reformulation of the usual Fernique–Talagrand theorem: since the cutoff $4\sigma(x)$ is not the standard full-radius presentation, the global radius $R_*$ must remain part of the anchored scale.

**Proposition 1.2** (Classical scale in anchored local form)**.** *There exist absolute constants $0 < c < C < \infty$ such that*

$$c\left(R_* + \inf_{\mu\in\Delta(T)} \Phi_{\mu,*}\right) \le \mathbb{E}\sup_{x\in T} X_x \le C\left(R_* + \inf_{\mu\in\Delta(T)} \Phi_{\mu,*}\right). \tag{1.6}$$

*Moreover, since the process is centered and anchored,*

$$\mathbb{E}\sup_{x\in T}|X_x| \asymp \mathbb{E}\sup_{x\in T} X_x.$$

*Therefore*

$$\mathbb{E}\sup_{x\in T}|X_x| \asymp R_* + \inf_{\mu\in\Delta(T)} \Phi_{\mu,*}. \tag{1.7}$$

We now record the in-expectation consequence of Theorem 1.1. In particular, after optimizing over the ambient prior, the resulting global in-expectation bound is sharp up to universal constants. The variance-radius term is not an additional loss: it is part of the anchored generic-chaining scale in (1.7), while the peeling logarithms disappear by taking $r_0 \ge \Phi_{\mu,*}$ and $s_0 = R_*$.

**Corollary 1.3** (In-expectation consequence and sharp global scale)**.** *Under the assumptions of Theorem 1.1, there exists an absolute constant $C > 0$ such that*

$$\mathbb{E}\left[\sup_{x\in T}\frac{\big(|X_x| - C\max\{\Phi_\mu(x), r_0\}\big)_+}{\max\{\sigma(x), s_0\}}\right] \le C\sqrt{\log(eM_{r_0,s_0})}. \tag{1.8}$$

*Consequently, for every $x \in T$,*

$$\mathbb{E}|X_x| \le C\max\{\Phi_\mu(x), r_0\} + C\max\{\sigma(x), s_0\}\sqrt{\log(eM_{r_0,s_0})}. \tag{1.9}$$

*At the global level, taking $r_0 \ge \Phi_{\mu,*}$ and $s_0 = R_*$ gives $M_{r_0,s_0} = 1$, hence*

$$\mathbb{E}\sup_{x\in T}|X_x| \le C(r_0 + R_*). \tag{1.10}$$

*Letting $r_0 \downarrow \Phi_{\mu,*}$, then optimizing over $\mu$, and using (1.7), we obtain the sharp equivalence*

$$\mathbb{E}\sup_{x\in T}|X_x| \asymp R_* + \inf_{\mu\in\Delta(T)}\Phi_{\mu,*}. \tag{1.11}$$

*Thus the optimized expectation version is optimal up to absolute constants.*

**Remark (variance term versus chaining complexity).** The term in Theorem 1.1 involving $\sigma(x)$ is a variance-sensitive Gaussian tail term. It is necessary for high-probability control, already for a single non-anchor point $X_x \sim N(0, \sigma(x)^2)$, but it is not a replacement for the chaining complexity $\Phi_\mu(x)$, which carries the local geometry of the index set. For a fixed marginal, $\mathbb{E}|X_x| = \sqrt{\frac{2}{\pi}}\,\sigma(x)$, so the extra factor $\sqrt{\log(eM_{r_0,s_0})}$ in (1.9) is not intrinsic to $\mathbb{E}|X_x|$. It is the price of extracting a simultaneous pointwise envelope by peeling first over the pointwise Fernique–Talagrand complexity and then over the local variance. At the global in-expectation level, this peeling factor disappears, yielding the sharp scale in (1.11).

**Remark (pointwise high probability).** For Gaussian processes, subset-level in-expectation bounds transfer to high-probability bounds through Borell–Tsirelson concentration, with variance scale $\sigma_H = \sup_{x\in H}(\mathbb{E}X_x^2)^{1/2}$ for a fixed subset $H$. The genuinely pointwise statement is not obtained from this transfer alone: after the subset-homogeneous estimate is established, one first applies uniform pointwise peeling to localize the Fernique–Talagrand term $\Phi_\mu(x)$, and then applies the same peeling mechanism a second time to improve the tail scale from the global radius to the local variance $\sigma(x)$. Conversely, the resulting high-probability envelope implies the stated in-expectation consequence by integrating the tail.

## 1.2 Main Bayesian algorithmic lower and upper bounds

The pointwise upper envelope is non-decision-theoretic, simultaneous, multiscale, and pathwise in the Gaussian field. A lower-bound analogue should therefore be stated with care. We first record a one-scale Bayesian and estimator-dependent primitive. Let $(T, d)$ be a finite metric space, let $\pi \in \Delta(T)$ be a known prior, and let $\{P_t : t \in T\}$ be an experiment on an observation space $\mathcal{Y}$. For an estimator $\widehat{t} : \mathcal{Y} \to T$, a reference law $Q \in \Delta(\mathcal{Y})$, and a scale $\Delta > 0$, define

$$\rho_{\Delta,Q}(\widehat{t}) := \mathbb{P}_{t\sim\pi,\, Y\sim Q}\{d(t, \widehat{t}(Y)) < \Delta\} = \mathbb{E}_{Y\sim Q}\,\pi\big(B_d(\widehat{t}(Y), \Delta)\big)\,,$$

and

$$\mathcal{I}_\pi(Q) := \mathbb{E}_{t\sim\pi} D_{\mathrm{KL}}(P_t \| Q).$$

The following result is the Gaussian-process and KL specialization of the interactive Fano method of Chen et al. (2024, Theorem 2). The original theorem provides a unifying information-theoretic principle for proving lower bounds in learning and decision-making problems.

**Proposition 1.4** (Bayesian algorithmic lower envelope)**.** *For every estimator $\widehat{t}$, every $\Delta > 0$, and every reference law $Q$ with $0 < \rho_{\Delta,Q}(\widehat{t}) < 1$,*

$$\mathbb{P}_{t\sim\pi,\, Y\sim P_t}\{d(t, \widehat{t}(Y)) \geq \Delta\} \geq \left(1 - \frac{\mathcal{I}_\pi(Q) + \log 2}{\log(1/\rho_{\Delta,Q}(\widehat{t}))}\right)_+. \tag{1.12}$$

*Consequently,*

$$\mathbb{E}_{t\sim\pi,\, Y\sim P_t} d(t, \widehat{t}(Y)) \geq \sup_{\substack{\Delta>0,\, Q:\\ 0<\rho_{\Delta,Q}(\widehat{t})<1}} \Delta \left(1 - \frac{\mathcal{I}_\pi(Q) + \log 2}{\log(1/\rho_{\Delta,Q}(\widehat{t}))}\right)_+. \tag{1.13}$$

The key point is that $\rho_{\Delta,Q}(\widehat{t})$ is not immediately replaced by the class-wide quantity

$$q_\pi(\Delta) := \sup_{a\in T} \pi(B_d(a, \Delta)).$$

Keeping the exact ghost mass $\mathbb{E}_{Y\sim Q}\pi(B_d(\widehat{t}(Y), \Delta))$ makes the lower bound algorithmic: it is attached to the image of the actual estimator under ghost data. Replacing it by $q_\pi(\Delta)$ gives the class-wide small-ball term used in classical Fano-type lower bounds (Zhang, 2006) and in the fractional-covering specialization of Chen et al. (2024). In the Gaussian nearest-neighbor comparison below, this small-ball term must still be balanced against the information radius of the Gaussian channel.

We next present a lower-bound comparison for the Gaussian process itself, rather than only for its induced distance. The result is stated for a comparison decoder. This includes ordinary nearest neighbor and the norm-regularized rule used in the separating example in Theorem 1.7.

**Theorem 1.5** (Algorithmic lower bound for comparison decoders)**.** *Let $v : T \to \mathcal{H}$ be an embedding into a finite-dimensional Hilbert space and set $d(s,t) = \|v_s - v_t\|_{\mathcal{H}}$. Let*

$$Y = v_\Theta + \tau Z, \qquad \Theta \sim \pi, \qquad Z \sim N(0, I_{\mathcal{H}}),$$

*and let $m \in \mathcal{H}$ be any reference center. Set*

$$Q_{\tau,m} := N(m, \tau^2 I_{\mathcal{H}}), \qquad \mathcal{I}_{\pi,m} := \frac{1}{2\tau^2}\int_T \|v_t - m\|^2\, \pi(dt).$$

*With $\bar{v}_\pi := \int_T v_t\, \pi(dt)$, the centered choice $m = \bar{v}_\pi$ gives*

$$\mathcal{I}_{\pi,\bar{v}_\pi} = \frac{\mathsf{V}_\pi}{4\tau^2}, \qquad \mathsf{V}_\pi := \iint d(s,t)^2 \pi(ds)\pi(dt).$$

*Let $\widehat{\Theta}(Y)$ be an estimator with values in $T$. Assume that there is a penalty $\Omega : T \to \mathbb{R}$ such that, under the true joint law,*

$$\|Y - v_{\widehat{\Theta}}\|^2 + \Omega(\widehat{\Theta}) \leq \|Y - v_\Theta\|^2 + \Omega(\Theta), \qquad \Omega(\widehat{\Theta}) \geq \Omega(\Theta) \quad a.s. \tag{1.14}$$

*If, for some* $c \in (0,1)$ *and* $\Delta > 0$*,*

$$\mathcal{I}_{\pi,m} + \log 2 \le (1-c)\log \frac{1}{\rho_{\Delta,Q_{\tau,m}}(\widehat{\Theta})}, \tag{1.15}$$

*then*

$$\mathbb{E}_{\Theta,Y}\, d(\Theta, \widehat{\Theta}(Y)) \ge c\Delta, \tag{1.16}$$

*and, for the Gaussian process* $X_t := \langle Z, v_t\rangle$ *built from the same Gaussian vector* $Z$*,*

$$\mathbb{E}_{\Theta,Z}\big[X_{\widehat{\Theta}} - X_\Theta\big] \ge \frac{c^2\Delta^2}{2\tau}. \tag{1.17}$$

*Thus the exact ghost mass gives an algorithmic lower bound on the Gaussian range sampled by the comparison decoder.*

*Ordinary nearest neighbor corresponds to* $\Omega \equiv 0$*. The unrestricted norm-regularized nearest-neighbor rule*

$$\widehat{\Theta}_\lambda(Y) \in \arg\min_{a\in T}\big\{\|Y - v_a\|^2 + \lambda\|v_a\|^2\big\}$$

*is also covered whenever the realized outputs have norm at least the norm of the true parameter, as in the separating example in Theorem 1.7. This is a canonical minimum-complexity inductive bias consistent with modern perspectives, rather than an explicit restriction of the hypothesis class.*

We defer to Appendix C a route from the Bayesian algorithmic lower bound to a Sudakov-type comparison via supremum relaxation, which consists of several technical results. Recent concurrent work (Zadik, 2026) shows that comparing nearest-neighbor and Bayes-optimal risks yields what is among the most succinct proofs to date of the majorizing-measure lower bound (van Handel, 2018; Zadik, 2026). Our aim here is instead conceptual: to show that the global Sudakov scale can be viewed as a coarse relaxation of a more localized algorithmic certificate.

The same pathwise comparison underlying Theorem 1.5, together with the pointwise Gaussian upper envelope, gives an algorithmic upper bound for Gaussian location decoding. Thus, when the two estimates match, the Bayesian algorithmic lower bound and the pointwise upper envelope certify the sharpness of one another.

**Corollary 1.6** (Algorithmic pointwise upper bound for estimator decoding)**.** *Work in the Gaussian location setup of Theorem 1.5. Thus* $T$ *is embedded as* $\{v_t : t \in T\}$ *in a Hilbert space,*

$$d(s,t) = \|v_s - v_t\|, \qquad Y = v_\Theta + \tau Z, \qquad X_t := \langle Z, v_t\rangle,$$

*where* $Z$ *is standard Gaussian. Let* $\widehat{\Theta} = \widehat{\Theta}(Y)$ *be a decoder satisfying the comparison condition* (1.14)*; in particular, the condition implies*

$$\|Y - v_{\widehat{\Theta}(Y)}\|^2 \le \|Y - v_\Theta\|^2 \tag{1.18}$$

*under the joint law of* $(\Theta, Y)$*.*

*Suppose that a deterministic envelope* $A : T \to [0,\infty)$ *satisfies*

$$\mathbb{P}_Z\big\{\forall t \in T:\ |X_t| \le A(t)\big\} \ge 1-\delta. \tag{1.19}$$

*Then, with probability at least* $1-\delta$ *under the joint law of* $(\Theta, Y)$*,*

$$d(\Theta, \widehat{\Theta}(Y))^2 \le 2\tau\big(A(\Theta) + A(\widehat{\Theta}(Y))\big). \tag{1.20}$$

*Consequently, if*

$$D_T := \sup_{s,t\in T} d(s,t) < \infty,$$

*then*

$$\mathbb{E}d(\Theta, \widehat{\Theta}(Y)) \le \mathbb{E}\sqrt{2\tau\big(A(\Theta) + A(\widehat{\Theta}(Y))\big)} + \delta D_T. \tag{1.21}$$

*In particular, after adjoining an anchor point* $t_0$ *with* $v_{t_0} = 0$ *and* $X_{t_0} = 0$ *if necessary, one may take*

$$A(t) = A_{\mu,r_0,s_0}(t;\delta)$$

*from* (1.5) *in Theorem 1.1, with the prior extended to* $T \cup \{t_0\}$*, for instance as* $\bar{\mu} = \frac{1}{2}\delta_{t_0} + \frac{1}{2}\mu$*, which changes the pointwise functional on* $T$ *only by universal constants. This yields a pointwise-complexity upper bound for the estimation risk.*

**Why keep the Bayesian and algorithmic form.** The Bayesian algorithmic form is crucial because a full-class minimax benchmark can be the wrong certificate for an overparameterized decision problem, while a restricted minimax benchmark can be *oracle-dependent*. Full minimax over $T$ can be too pessimistic, because it is governed by the worst point of the ambient class rather than by the geometry selected by the prior and the algorithm; maximizing over priors can similarly erase the structure that makes a particular estimator appropriate. On the other hand, a restricted minimax benchmark is useful only when the correct restricted class is known in advance. In the separating example in Theorem 1.7, the relevant class is the hub–leaf subatlas $H_M$, but identifying $H_M$ as the right estimator-selected subatlas is precisely the decision-dependent information revealed by the ghost image $Y \mapsto \widehat{t}(Y)$. Along the sequence $N = M^K$ with $K \to \infty$, the norm-regularized decoder uses a fixed numerical penalty $\lambda$, independent of the prior $\pi$ and the problem parameters $M, N, R, S, \tau$, and has Bayes risk $\asymp R$, whereas the full-class minimax risk is $\asymp S$, with $\frac{S}{R} \asymp \sqrt{\frac{\log N}{\log M}} = \sqrt{K} \to \infty$. Thus the full-class benchmark misses the decision-aligned scale, while the restricted benchmark recovers it only after an oracle has selected $H_M$.

The same issue is compounded in deep networks: the effective subatlas is rarely a fixed combinatorial class and may depend on the architecture, data distribution, learned representation, and compressed feature spectrum (Li and Xu, 2026). Moreover, computationally efficient procedures may also operate through relaxations or implicit regularization rather than by solving an exact restricted-class problem, as in sparse regression. The algorithmic ghost mass $\rho_{\Delta,Q}(\widehat{t})$ is designed for this regime: it evaluates the estimator actually used, under the intended prior geometry, and therefore can certify a pointwise upper envelope without requiring a separate restricted class to be specified in advance.

The Bayesian *algorithmic* lower bounds in Proposition 1.4 and Theorem 1.5 should also not be read as ordinary Bayesian lower bounds whose purpose is to optimize over all estimators given a specific prior. The prior is an averaging and comparison device, while the certified object is the performance of a concrete estimator that is of practical interest regardless of knowledge of the prior. This distinction is important because the usual Fano relaxation may be vacuous for a fixed nonuniform prior even though the estimator of interest remains hard on the subgeometry selected by the prior and the ghost law. Such vacuity does not mean that the Bayesian decision problem is easy; it means only that the particular worst-case small-ball relaxation $\sup_a \pi(B_d(a,\Delta))$ has discarded the algorithmic geometry of the estimator. In the separating example of Theorem 1.7, the Bayes-optimal risk under the prior is also $\asymp R$, even though the Bayes oracle can be highly non-analytic. The same rotationally symmetric class construction can accommodate an arbitrarily large finite family of priors that exhibit the separation and have markedly different local mass profiles: the Bayes-optimal oracle varies substantially across these priors, whereas the unrestricted estimator remains unchanged.

**Relationship to the pointwise upper theorem and existing lower bounds.** The Bayesian algorithmic lower bound also clarifies the relation to the pointwise upper theorem through Corollary 1.6. The upper envelope is simultaneous, multiscale, and high-probability: it is pathwise in the Gaussian field and integrates local prior masses across scales. By contrast, the lower bound is Bayesian, one-scale, and estimator-dependent: it averages over $\Theta \sim \pi$ and $Y \sim P_\Theta$ for a fixed estimator, while the additional ghost/reference law $Y \sim Q$ enters through the quantile term $\rho_{\Delta,Q}(\widehat{t})$. This parallels the information-theoretic upper–lower duality emphasized by Zhang (2006), but with the additional point that the lower bound is a data-processing statement valid for every estimator.

DEC-type lower bounds (Foster et al., 2021, 2023) provide an independent motivation for incorporating algorithmic dependence into the fundamental limits of online regret, exploration, and interactive decision making. This perspective further motivates the Bayesian algorithmic lower bound of Chen et al. (2024) and related variants (Xu and Zeevi, 2025b). The present algorithmic lower-bound perspective isolates a distinct but complementary principle: a prior-mass lower-bound analogue of pointwise complexity, in the spirit of the pointwise generalization framework of Li and Xu (2026) for deep representation learning.

## 1.3 Separating example construction and pointwise certification

The next theorem is the formal version of the separating example proved in Section 4. It shows that, in an overparameterized ambient class, a fixed-prior worst-case Fano relaxation, a Bayesian algorithmic lower envelope, a pointwise Gaussian upper envelope, and the full minimax/global Gaussian scale can all have different meanings and different orders. The estimator is unrestricted over the full ambient class; the relevant subatlas is selected by a norm-regularized inductive bias and by the ghost law, rather than imposed as a constraint.

**Theorem 1.7** (Decision-aligned separation and pointwise certification)**.** *There exist universal constants $0 < c < C < \infty$ with the following property. Let $M, N \geq 2$ satisfy $\log N \geq 1024 \log M$. Let*

$$e_0, e_1, \ldots, e_M, f_1, \ldots, f_N$$

*be orthonormal vectors in $\mathbb{R}^{M+N+1}$. Fix $R > 0$, define*

$$\tau := \frac{R}{\sqrt{\log M}}, \qquad S := \frac{1}{8}\tau\sqrt{\log N} = \frac{R}{8}\sqrt{\frac{\log N}{\log M}},$$

*and set*

$$H_M := \{Re_0, Re_1, \ldots, Re_M\}, \qquad C_N := \{Sf_1, \ldots, Sf_N\}, \qquad T_{M,N} := H_M \cup C_N.$$

*Let $d(s,t) = \|s - t\|_2$, let $\Delta = R/3$, and let the prior be*

$$\pi(Re_0) = \frac{1}{2}, \qquad \pi(Re_i) = \frac{1}{2M}, \quad i = 1, \ldots, M, \qquad \pi(Sf_j) = 0, \quad j = 1, \ldots, N.$$

*Consider the Gaussian location experiment*

$$Y = \Theta + \tau Z, \qquad \Theta \sim \pi, \qquad Z \sim N(0, I_{M+N+1}),$$

*and the unrestricted norm-regularized decoder*

$$\widehat{\Theta}_\lambda(Y) \in \arg\min_{a\in T_{M,N}} \{\|Y-a\|^2 + \lambda\|a\|^2\}, \qquad \lambda = 64. \tag{1.22}$$

*Let $Q_\tau = N(0, \tau^2 I_{M+N+1})$. Then, for all sufficiently large $M, N$, the following hold.*

(i) Fixed-prior Fano relaxation is vacuous. *Since every $\Delta$-ball in $T_{M,N}$ is a singleton,*

$$q_\pi(\Delta) := \sup_{a\in T_{M,N}} \pi(B_d(a,\Delta)) = \frac{1}{2}.$$

*Moreover the usual data-processing lower envelope obtained by replacing the exact ghost mass by $q_\pi(\Delta)$ gives zero at the reference law $Q_\tau$.*

(ii) Exact algorithmic ghost mass is small. *The estimator-dependent ghost mass satisfies*

$$\rho_{\Delta,Q_\tau}(\widehat{\Theta}_\lambda) = \mathbb{E}_{Y\sim Q_\tau}\pi(B_d(\widehat{\Theta}_\lambda(Y),\Delta)) \le \frac{C}{M}.$$

*Consequently the exact ghost entropy is of order* $\log M$*, whereas the worst-case Fano small-ball entropy for the same prior is only* $\log 2$.

(iii) Sharp Bayes risk for the actual unrestricted estimator.

$$cR \le \mathbb{E}_{\Theta,Y} d(\Theta, \widehat{\Theta}_\lambda(Y)) \le CR.$$

*The lower bound follows from Proposition 1.4 using the exact ghost mass; the upper bound is the actual Bayes risk of the same unrestricted estimator.*

*Moreover, the Bayes-optimal risk is also $\asymp R$, despite the oracle estimator being highly non-analytic. Thus the class-wide Fano relaxation below is vacuous only as a certificate; it does not reflect an easy Bayes problem. The same rotationally symmetric class construction can accommodate an arbitrarily large finite family of priors exhibiting the separation, with markedly different local mass profiles. Across this family, the Bayes-optimal oracle may vary substantially, while the unrestricted estimator remains unchanged.*

(iv) Full-class minimax risk is much larger. *If*

$$\mathfrak{R}_{\mathrm{mm}}(T_{M,N},\tau) := \inf_{\widehat{\theta}} \sup_{t\in T_{M,N}} \mathbb{E}_t d(t, \widehat{\theta}(Y)),$$

*where the infimum is over all measurable estimators with values in the ambient Euclidean space, then*

$$cS \le \mathfrak{R}_{\mathrm{mm}}(T_{M,N},\tau) \le CS.$$

*Thus, whenever* $\log N/\log M \to \infty$*, the full-class minimax risk is asymptotically larger than the Bayes risk under $\pi$.*

(v) Algorithmically induced Gaussian certificate matches the selected subatlas. *For the Gaussian process*

$$X_t := \langle Z, t\rangle, \qquad t \in T_{M,N},$$

*where $Z$ is the same standard Gaussian vector as in the location experiment $Y = \Theta + \tau Z$, define*

$$\mathcal{G}_{\mathrm{alg}} := \mathbb{E}_{\Theta,Z}\big[X_{\widehat{\Theta}_\lambda(\Theta+\tau Z)} - X_\Theta\big].$$

*Then*

$$\mathcal{G}_{\mathrm{alg}} \asymp \mathbb{E} \sup_{s,u \in H_M} (X_s - X_u) \asymp R\sqrt{\log M}.$$

*Moreover, under the ghost/reference law* $Y \sim Q_\tau = N(0, \tau^2 I_{M+N+1})$,

$$Q_\tau\{\widehat{\Theta}_\lambda(Y) \in C_N\} \le N^{-2}$$

*for all sufficiently large* $N$*, and hence the ghost-selected subatlas is* $H_M$ *up to negligible probability. Conditional on* $\widehat{\Theta}_\lambda(Y) \in H_M$*, the ghost selection law is uniform on* $H_M$*. Therefore, after adjoining the zero anchor, the pointwise Gaussian upper theorem applied to the restricted process on* $H_M$ *gives the matching high-probability envelope*

$$\mathbb{P}\left(\forall t \in H_M : \ |X_t| \le CR\sqrt{\log \frac{M}{\delta}}\right) \ge 1 - \delta.$$

(vi) Global Gaussian scale can be arbitrarily larger.

$$\mathbb{E} \sup_{s,u \in T_{M,N}} (X_s - X_u) \ge cS\sqrt{\log N} = \frac{c}{8} R \frac{\log N}{\sqrt{\log M}}.$$

*Thus the global Gaussian scale exceeds the decision-aligned scale* $R\sqrt{\log M}$ *by a factor of order* $\log N / \log M$*, which can be made arbitrarily large by taking* $N = M^K$ *with* $K \to \infty$*.*

*Consequently the same explicit problem sequence separates*

$$\begin{aligned} & \textit{fixed-prior Fano relaxation} \\ \ll\ & \textit{algorithmic ghost-mass lower envelope} \\ \asymp\ & \textit{pointwise upper envelope on the selected subatlas} \\ \ll\ & \textit{full-class minimax/global Gaussian scale.} \end{aligned}$$

### 1.4 Organization

Section 2 proves the pointwise Gaussian majorizing-measure theorem. Section 3 proves the Bayesian algorithmic lower and upper bounds. Section 4 proves the separating example and pointwise certification. Section 5 closes the main text. Appendix A contains the finite-cutoff renormalization material. Appendix B contains the graph local-time application and cover-time comparison. Appendix C records technical variants and Sudakov-type consequences.

The appendices are intentionally separated from the main Gaussian and Bayesian results. They record structural analogies and first finite examples, rather than claiming a complete renormalization group theory or a complete theory of local time, cover time, and blanket time.

## 2 Proof of Pointwise Gaussian Majorizing-Measure Theorem

The proof follows a subset-homogeneity and peeling strategy. First, for each fixed subset $H \subseteq T$, we combine an anchored form of Fernique's majorizing-measure upper bound (Fernique, 1975; Talagrand, 1987), the ambient equivalence principle of pointwise dimension, and Borell–Tsirelson concentration. Second, we apply the one-parameter uniform pointwise peeling lemma twice: first to localize the Fernique–Talagrand integral $\Phi_\mu(x)$, and then to localize the variance scale from $R_*$ to $\sigma(x)$. This avoids introducing a separate two-dimensional peeling corollary.

The following uniform pointwise peeling lemma was originally developed for empirical processes in Xu and Zeevi (2025a). It identifies minimal conditions for obtaining pointwise envelopes, up to negligible double-logarithmic factors, without imposing unrealistic Bernstein-type or rigid sub-root assumptions. Necessary and sufficient justifications and discussions of data dependence are provided in Li and Xu (2026). Here we extend the lemma to general indexed processes, including Gaussian processes.

**Lemma 2.1** (Uniform pointwise peeling for indexed processes)**.** *Let $\{Z_x\}_{x\in T}$ be random variables, let $d_0 : T \to [0, R]$, and let $\psi(r;\delta)$ be nondecreasing in $r$ and nonincreasing in $\delta$. Assume that for every fixed subset $H \subseteq T$ and every $\delta \in (0,1)$,*

$$\mathbb{P}\left(\sup_{x\in H} Z_x \le \sup_{x\in H} \psi(d_0(x);\delta)\right) \ge 1-\delta. \tag{2.1}$$

*Then for every $r_0 > 0$, with probability at least $1-\delta$, uniformly over all $x \in T$,*

$$Z_x \le \psi\left(\max\{2d_0(x), r_0\}; \frac{\delta}{1+\lceil \log_2^+(R/r_0)\rceil}\right), \tag{2.2}$$

*where* $\log_2^+(u) = \max\{\log_2 u, 0\}$.

*Proof.* Let $m := \lceil \log_2^+(R/r_0)\rceil$. The sets

$$H_0 := \{x : d_0(x) \le r_0\}, \qquad H_j := \{x : 2^{j-1}r_0 < d_0(x) \le 2^j r_0\} \quad (1 \le j \le m)$$

cover $T$. Apply (2.1) to each $H_j$ with confidence level $\delta/(m+1)$, and take a union bound. On the resulting event, if $x \in H_0$, then

$$Z_x \le \psi\left(r_0; \frac{\delta}{m+1}\right) \le \psi\left(\max\{2d_0(x), r_0\}; \frac{\delta}{m+1}\right),$$

and if $x \in H_j$ for $j \ge 1$, then $2^j r_0 \le 2d_0(x)$, so

$$Z_x \le \psi\left(2^j r_0; \frac{\delta}{m+1}\right) \le \psi\left(\max\{2d_0(x), r_0\}; \frac{\delta}{m+1}\right).$$

This proves the claim. □

We also use the following localized version of the ambient equivalence principle of pointwise dimension from Lemma 7 of Li and Xu (2026).

**Lemma 2.2** (Ambient equivalence for the localized integral)**.** *Let $(T, d)$ be a finite metric space and let $H \subseteq T$. Let $p : T \to H$ be a nearest-point selector, so that $d(y, p(y)) = \min_{h\in H} d(y, h)$, and let $\mu_H = p_\#\mu$ be the pushforward of an ambient prior $\mu \in \Delta(T)$. Then, for every $x \in H$ and every $\varepsilon > 0$,*

$$\mu_H(B_d(x, 2\varepsilon)) \ge \mu(B_d(x, \varepsilon)).$$

*Consequently, for the localized functional $\Phi$ in (1.1),*

$$\Phi_{\mu_H}(x) \le 2\Phi_\mu(x), \qquad x \in H. \tag{2.3}$$

*Proof.* If $y \in B_d(x, \varepsilon)$, then

$$d(p(y), x) \le d(p(y), y) + d(y, x) \le 2d(y, x) \le 2\varepsilon,$$

because $x \in H$ and $p(y)$ is a nearest point in $H$. Hence $p(B_d(x, \varepsilon)) \subseteq B_d(x, 2\varepsilon)$, which gives the ball-mass inequality. Therefore

$$\Phi_{\mu_H}(x) = \int_0^{4\sigma(x)} \sqrt{\log \frac{1}{\mu_H(B_d(x,u))}}\, du \le 2 \int_0^{2\sigma(x)} \sqrt{\log \frac{1}{\mu(B_d(x,\varepsilon))}}\, d\varepsilon \le 2\Phi_\mu(x),$$

where we used the change of variables $u = 2\varepsilon$. □

**Lemma 2.3** (Anchored subset majorizing-measure bound)**.** *For every fixed subset $H \subseteq T$, every ambient prior $\mu \in \Delta(T)$, and*

$$\sigma_H := \sup_{x \in H} \sigma(x),$$

*there is an absolute constant $C > 0$ such that*

$$\mathbb{E} \sup_{x \in H} X_x \le C \sup_{x \in H} \Phi_\mu(x) + C\sigma_H. \tag{2.4}$$

*Proof.* Let $H_0 := H \cup \{x_0\}$, and let $p : T \to H_0$ be a nearest-point selector. Define the probability measure

$$\bar{\mu}_H := \frac{1}{2}\delta_{x_0} + \frac{1}{2} p_{\#}\mu$$

on $H_0$. Fernique's majorizing-measure upper bound (Fernique, 1975; Talagrand, 1987), applied to the restricted process on $H_0$, gives

$$\mathbb{E} \sup_{x \in H} X_x \le C \sup_{x \in H_0} \int_0^{2\sigma_H} \sqrt{\log \frac{1}{\bar{\mu}_H(B_d(x,u))}}\, du,$$

since $\operatorname{diam}(H_0, d) \le 2\sigma_H$. For $x = x_0$, the integral is at most $2\sigma_H\sqrt{\log 2}$, because $\bar{\mu}_H\{x_0\} \ge 1/2$. For $x \in H$, split the integral at $4\sigma(x) \wedge 2\sigma_H$. If $4\sigma(x) > 2\sigma_H$, the second interval below is empty. On $[4\sigma(x), 2\sigma_H]$, the ball $B_d(x, u)$ contains $x_0$, hence has $\bar{\mu}_H$-mass at least $1/2$, so this part is bounded by $C\sigma_H$. On $[0, 4\sigma(x)]$, Lemma 2.2 and the extra factor $1/2$ in $\bar{\mu}_H$ give a bound of order $\Phi_\mu(x) + \sigma(x)$. Taking the supremum over $x \in H$ proves (2.4). □

**Lemma 2.4** (Subset-homogeneous Gaussian estimate)**.** *For every fixed subset $H \subseteq T$, every ambient prior $\mu \in \Delta(T)$, and every $\delta \in (0, 1)$,*

$$\mathbb{P}\left( \sup_{x \in H} X_x \le C \sup_{x \in H} \Phi_\mu(x) + C\sigma_H \sqrt{\log(e/\delta)} \right) \ge 1 - \delta, \tag{2.5}$$

*where $\sigma_H = \sup_{x \in H} \sigma(x)$.*

*Proof.* By Lemma 2.3,

$$\mathbb{E} \sup_{x \in H} X_x \le C \sup_{x \in H} \Phi_\mu(x) + C\sigma_H.$$

Realize $X_x = \langle g, v_x \rangle$ as an isonormal process on a Hilbert space. The map $g \mapsto \sup_{x \in H}\langle g, v_x \rangle$ is $\sigma_H$-Lipschitz. Borell–Tsirelson concentration therefore gives the desired inequality, after absorbing the expectation term $C\sigma_H$ into $C\sigma_H\sqrt{\log(e/\delta)}$. □

*Proof of Theorem 1.1.* Fix an arbitrary subset $H \subseteq T$. Applying Lemma 2.1 on the index set $H$, with $Z_x = X_x$, $d_0(x) = \Phi_\mu(x)$, $R = \Phi_{\mu,*}$, and using the subset estimate (2.5), yields the following: for every $r_0 > 0$ and every $\delta \in (0,1)$, with probability at least $1 - \delta$,

$$\forall x \in H : \qquad X_x \le C \max\{\Phi_\mu(x), r_0\} + C\sigma_H \sqrt{\log\left(\frac{eM_\Phi}{\delta}\right)}, \tag{2.6}$$

where

$$M_\Phi := 1 + \lceil \log_2^+(\Phi_{\mu,*}/r_0) \rceil.$$

Since $H$ was arbitrary, (2.6) is now a subset-homogeneous estimate for the centered residual

$$Y_x := X_x - C \max\{\Phi_\mu(x), r_0\}$$

with complexity $\sigma(x)$. Applying Lemma 2.1 a second time, now with $Z_x = Y_x$, $d_0(x) = \sigma(x)$, $R = R_*$, and

$$\psi(s;\delta) := Cs\sqrt{\log\left(\frac{eM_\Phi}{\delta}\right)},$$

gives, with probability at least $1 - \delta$,

$$\forall x \in T : \qquad Y_x \le C \max\{\sigma(x), s_0\} \sqrt{\log\left(\frac{eM_\Phi M_\sigma}{\delta}\right)},$$

where

$$M_\sigma := 1 + \lceil \log_2(R_*/s_0) \rceil.$$

Combining the last display with the definition of $Y_x$ proves (1.4), since $M_{r_0,s_0} = M_\Phi M_\sigma$. The absolute-value estimate (1.5) follows by applying the same argument to $-X$ and taking a union bound. □

*Proof of Proposition 1.2.* The upper bound follows from Lemma 2.3 with $H = T$, followed by optimization over $\mu$. For the lower bound, choose $x_*$ with $\sigma(x_*) = R_*$. Since $X_{x_0} = 0$,

$$\sup_{x \in T} X_x \ge (X_{x_*})_+, \qquad \mathbb{E}(X_{x_*})_+ = \frac{R_*}{\sqrt{2\pi}}.$$

It remains to compare the optimized local functional with the classical full-radius one. Let

$$\Phi_\mu^{\mathrm{full}}(x) := \int_0^{2R_*} \sqrt{\log \frac{1}{\mu(B_d(x,\varepsilon))}}\, d\varepsilon.$$

The classical majorizing-measure theorem gives

$$\inf_{\mu \in \Delta(T)} \sup_x \Phi_\mu^{\mathrm{full}}(x) \lesssim \mathbb{E} \sup_{x \in T} X_x.$$

Since $\Phi_\mu(x) \le \Phi_\mu^{\mathrm{full}}(x)$ for every $x$, we have

$$\inf_\mu \Phi_{\mu,*} \le \inf_\mu \sup_x \Phi_\mu^{\mathrm{full}}(x) \lesssim \mathbb{E} \sup_{x \in T} X_x.$$

Together with the radius lower bound this proves the left side of (1.6). The absolute-value statement follows from

$$\sup_x X_x \le \sup_x |X_x| \le \sup_x X_x + \sup_x(-X_x)$$

and symmetry. □

*Proof of Corollary 1.3.* Let

$$Y := \sup_{x\in T} \frac{\big(|X_x| - C\max\{\Phi_\mu(x), r_0\}\big)_+}{\max\{\sigma(x), s_0\}}.$$

By the two-sided form of Theorem 1.1, after increasing $C$ if necessary,

$$\mathbb{P}\left(Y > C\sqrt{\log\left(\frac{eM_{r_0,s_0}}{\delta}\right)}\right) \le \delta, \qquad \delta \in (0,1).$$

Integrating this tail gives

$$\mathbb{E}Y \le C\sqrt{\log(eM_{r_0,s_0})},$$

which proves (1.8). The pointwise bound (1.9) follows by dropping the supremum, and the global bound (1.10) follows by taking $r_0 \ge \Phi_{\mu,*}$, $s_0 = R_*$, and then letting $r_0 \downarrow \Phi_{\mu,*}$. Optimizing over $\mu$ and applying Proposition 1.2 proves (1.11). □

## 3 Proof of Bayesian Algorithmic Lower and Upper Bounds

*Proof of Proposition 1.4.* The result is the Gaussian-process and KL specialization of the interactive Fano method of Chen et al. (2024, Theorem 2); we give a self-contained proof. Fix $\widehat{t}$, $Q$, and $\Delta$, and let

$$E := \{(t, Y) : d(t, \widehat{t}(Y)) < \Delta\}.$$

Under the true joint law $P := \int \pi(dt)P_t(dY)$, write $p = P(E)$. Under the ghost law $P_0 := \pi \otimes Q$, write

$$q = P_0(E) = \rho_{\Delta,Q}(\widehat{t}).$$

By data processing,

$$\mathcal{I}_\pi(Q) = D_{\mathrm{KL}}(P\|P_0) \ge D_{\mathrm{KL}}(\mathrm{Bern}(p)\|\,\mathrm{Bern}(q)).$$

The elementary inequality

$$D_{\mathrm{KL}}(\mathrm{Bern}(p)\|\,\mathrm{Bern}(q)) \ge p\log\frac{1}{q} - \log 2$$

gives

$$p \le \frac{\mathcal{I}_\pi(Q) + \log 2}{\log(1/q)}.$$

Therefore

$$P\{d(t, \widehat{t}(Y)) \ge \Delta\} = 1 - p \ge 1 - \frac{\mathcal{I}_\pi(Q) + \log 2}{\log(1/\rho_{\Delta,Q}(\widehat{t}))},$$

with the positive part inserted to make the bound nonnegative. Multiplying by $\Delta$ and optimizing over $\Delta, Q$ proves (1.13). □

*Proof of Theorem 1.5.* For the Gaussian location experiment and the reference law $Q_{\tau,m}$,

$$\mathcal{I}_\pi(Q_{\tau,m}) = \int_T D_{\mathrm{KL}}\big(N(v_t, \tau^2 I)\|N(m, \tau^2 I)\big)\,\pi(dt) = \frac{1}{2\tau^2}\int_T \|v_t - m\|^2\pi(dt) = \mathcal{I}_{\pi,m}.$$

Applying Proposition 1.4 to $\widehat{\Theta}$ and $Q_{\tau,m}$ gives (1.16) under condition (1.15).

It remains to relate this Bayes location error to a Gaussian range. By the comparison-decoder condition (1.14),

$$\|Y - v_{\widehat{\Theta}}\|^2 - \|Y - v_\Theta\|^2 \le \Omega(\Theta) - \Omega(\widehat{\Theta}) \le 0.$$

Since $Y = v_\Theta + \tau Z$, expanding and cancelling $\tau^2\|Z\|^2$ gives

$$\|v_\Theta - v_{\widehat{\Theta}}\|^2 + 2\tau\langle Z, v_\Theta - v_{\widehat{\Theta}}\rangle \le 0,$$

and hence

$$\|v_\Theta - v_{\widehat{\Theta}}\|^2 \le 2\tau\langle Z, v_{\widehat{\Theta}} - v_\Theta\rangle = 2\tau\big(X_{\widehat{\Theta}} - X_\Theta\big).$$

Taking expectations and using Jensen's inequality,

$$c^2\Delta^2 \le \Big(\mathbb{E}d(\Theta, \widehat{\Theta})\Big)^2 \le \mathbb{E}d(\Theta, \widehat{\Theta})^2 \le 2\tau\,\mathbb{E}\big[X_{\widehat{\Theta}} - X_\Theta\big].$$

This proves (1.17). □

*Proof of Corollary 1.6.* By the comparison condition (1.18) and the identity $Y = v_\Theta + \tau Z$,

$$\|v_\Theta + \tau Z - v_{\widehat{\Theta}}\|^2 \le \|\tau Z\|^2.$$

Expanding the left-hand side and cancelling the common term $\tau^2\|Z\|^2$, we obtain the pathwise inequality

$$d(\Theta, \widehat{\Theta})^2 = \|v_\Theta - v_{\widehat{\Theta}}\|^2 \le 2\tau\langle Z, v_{\widehat{\Theta}} - v_\Theta\rangle = 2\tau\big(X_{\widehat{\Theta}} - X_\Theta\big).$$

On the event in (1.19), which holds with probability at least $1 - \delta$ and is uniform over all $t \in T$,

$$X_{\widehat{\Theta}} - X_\Theta \le |X_{\widehat{\Theta}}| + |X_\Theta| \le A(\widehat{\Theta}) + A(\Theta).$$

Hence, on the same event,

$$d(\Theta, \widehat{\Theta})^2 \le 2\tau\big(A(\Theta) + A(\widehat{\Theta})\big),$$

which proves (1.20).

For the expectation bound, let $E$ denote the event $\{\forall t \in T:\ |X_t| \le A(t)\}$. On $E$, the preceding display gives

$$d(\Theta, \widehat{\Theta}) \le \sqrt{2\tau\big(A(\Theta) + A(\widehat{\Theta})\big)}.$$

On $E^c$, we use the trivial bound

$$d(\Theta, \widehat{\Theta}) \le D_T.$$

Therefore

$$\begin{aligned}\mathbb{E}d(\Theta, \widehat{\Theta}) &\le \mathbb{E}\left[\sqrt{2\tau\big(A(\Theta) + A(\widehat{\Theta})\big)}\,\mathbf{1}_E\right] + D_T\,\mathbb{P}(E^c) \\ &\le \mathbb{E}\sqrt{2\tau\big(A(\Theta) + A(\widehat{\Theta})\big)} + \delta D_T,\end{aligned}$$

which proves (1.21). The final claim follows by applying the absolute pointwise Gaussian envelope (1.5) to the anchored process $(X_t)_{t\in T\cup\{t_0\}}$ and then restricting back to $T$. □

## 4 Proof for Separating Example and Pointwise Certification

This section proves Theorem 1.7. The example is a finite-dimensional weighted-basis construction. It is intentionally elementary, but it captures the benchmark separation needed for the paper: the full ambient class is overparameterized, the correct subatlas is decision-dependent, and the actual estimator is unrestricted over the full class but uses a minimum-norm inductive bias. The proof keeps the discussion in separate paragraphs because each paragraph isolates one benchmark in the separation.

*Proof of Theorem 1.7.* Let $M, N \geq 2$ satisfy $\log N \geq 1024 \log M$. Let

$$e_0, e_1, \ldots, e_M, f_1, \ldots, f_N$$

be orthonormal vectors in $\mathbb{R}^{M+N+1}$, and define

$$\tau := \frac{R}{\sqrt{\log M}}, \qquad S := \frac{1}{8}\tau\sqrt{\log N} = \frac{R}{8}\sqrt{\frac{\log N}{\log M}}.$$

The assumption on $M, N$ implies $S \geq 4R$. Set

$$H_M := \{Re_0, Re_1, \ldots, Re_M\}, \qquad C_N := \{Sf_1, \ldots, Sf_N\}, \qquad T_{M,N} := H_M \cup C_N,$$

with Euclidean metric $d(s,t) = \|s - t\|_2$. The prior is

$$\pi(Re_0) = \frac{1}{2}, \qquad \pi(Re_i) = \frac{1}{2M}, \quad i = 1, \ldots, M, \qquad \pi(Sf_j) = 0, \quad j = 1, \ldots, N,$$

and we take $\Delta = R/3$. Since $S \geq 4R$, every $\Delta$-ball in $T_{M,N}$ is a singleton. Hence

$$q_\pi(\Delta) = \sup_{a \in T_{M,N}} \pi(B_d(a, \Delta)) = \frac{1}{2}.$$

This proves the fixed-prior small-ball claim in part (i).

**Algorithmic ghost mass versus worst-case Fano mass.** Let

$$Y = \Theta + \tau Z, \qquad \Theta \sim \pi, \qquad Z \sim N(0, I_{M+N+1}),$$

and use the ghost/reference law $Q_\tau = N(0, \tau^2 I_{M+N+1})$. Under $Q_\tau$, the scores of the $M+1$ lower-norm points in $H_M$ are exchangeable. We first show that the cloud is selected with negligible ghost probability. If a cloud point $Sf_j$ beats the hub point $Re_0$, then

$$2\langle Y, Sf_j\rangle - (1+\lambda)S^2 \geq 2\langle Y, Re_0\rangle - (1+\lambda)R^2.$$

Under $Q_\tau$, the random part of the difference is centered Gaussian with variance at most $4\tau^2(S^2+R^2)$. Since $S \geq 4R$ and $\lambda = 64$, a Gaussian union bound gives

$$Q_\tau(\widehat{\Theta}_\lambda \in C_N) \leq N^{-2} \tag{4.1}$$

for all sufficiently large $N$. Conditional on $\widehat{\Theta}_\lambda \in H_M$, exchangeability gives

$$Q_\tau(\widehat{\Theta}_\lambda = Re_0 \mid \widehat{\Theta}_\lambda \in H_M) = \frac{1}{M+1}.$$

Consequently,

$$Q_\tau(\widehat{\Theta}_\lambda = Re_0) \le \frac{1}{M+1} + N^{-2}.$$

Since every $\Delta$-ball is a singleton and $\pi$ vanishes on the cloud,

$$\begin{aligned}\rho_{\Delta,Q_\tau}(\widehat{\Theta}_\lambda) &= \mathbb{E}_{Y\sim Q_\tau}\pi(B_d(\widehat{\Theta}_\lambda(Y),\Delta)) \\ &\le \frac{1}{2}Q_\tau(\widehat{\Theta}_\lambda = Re_0) + \frac{1}{2M}Q_\tau(\widehat{\Theta}_\lambda \in \{Re_1,\dots,Re_M\}) \le \frac{2}{M}\end{aligned}$$

for all large $M, N$. Hence

$$\log\frac{1}{\rho_{\Delta,Q_\tau}(\widehat{\Theta}_\lambda)} \ge \log M - O(1),$$

whereas the worst-case relaxation for the same prior gives only $\log 2$. This proves part (ii) and the entropy separation in part (i).

**Sharp algorithmic lower bound for the actual unrestricted estimator.** The information radius under $Q_\tau$ is

$$\mathsf{I}_{\pi,Q_\tau} = \mathbb{E}_{\Theta\sim\pi}D_{\mathrm{KL}}\big(N(\Theta,\tau^2 I)\,\|\,N(0,\tau^2 I)\big) = \frac{\mathbb{E}_\pi\|\Theta\|^2}{2\tau^2} = \frac{R^2}{2\tau^2} = \frac{1}{2}\log M.$$

Combining this with the ghost-mass bound, for all large $M$,

$$\mathsf{I}_{\pi,Q_\tau} + \log 2 \le \frac{3}{4}\log\frac{1}{\rho_{\Delta,Q_\tau}(\widehat{\Theta}_\lambda)}.$$

Proposition 1.4 therefore yields

$$\mathbb{E}_{\Theta,Y}d(\Theta,\widehat{\Theta}_\lambda(Y)) \ge c\Delta \asymp R. \tag{4.2}$$

This lower bound is sharp for the actual Bayesian decision problem. When $\Theta \in H_M$, all true parameters have norm $R$. The same score comparison as above, now under $Y = \Theta + \tau Z$, gives

$$\mathbb{P}_{\Theta,Y}(\widehat{\Theta}_\lambda \in C_N) \le N^{-2}.$$

On the complementary event, both $\Theta$ and $\widehat{\Theta}_\lambda$ lie in $H_M$, whose diameter is at most $\sqrt{2}R$. Hence

$$\mathbb{E}_{\Theta,Y}d(\Theta,\widehat{\Theta}_\lambda(Y)) \le \sqrt{2}R + N^{-2}S \lesssim R. \tag{4.3}$$

Combining the algorithmic lower bound (4.2) and the algorithmic upper bound (4.3) proves the Bayes risk statement for the estimator $\widehat{\Theta}_\lambda$:

$$E_{\Theta,Y}d(\Theta,\widehat{\Theta}_\lambda(Y)) \asymp R.$$

This proves the statement about the Bayes risk of the actual estimator in part (iii).

**Bayes-optimal risk is of the same order.** The Bayes-optimal risk for this prior is also $\asymp R$, despite the oracle estimator being highly non-analytic. The upper bound follows from (4.3). For the lower bound, condition on the event $\mathcal{L}$ that $\Theta$ is one of the leaves. Under this conditional experiment, write

$$J \sim \mathrm{Unif}\{1,\dots,M\}, \qquad \Theta = Re_J, \qquad Y = Re_J + \tau Z, \quad Z \sim N(0, I_{M+N+1}).$$

Given any estimator $\widehat{\theta}(Y)$, define the induced leaf classifier

$$\widehat{J}(Y) \in \arg\min_{1\le j\le M} \|\widehat{\theta}(Y) - Re_j\|,$$

with ties broken arbitrarily. Since the leaves are separated by $\sqrt{2}R$, the event $\|\widehat{\theta}(Y) - Re_J\| < R/2$ implies $\widehat{J} = J$. Hence

$$\mathbb{E}\Big[\|\widehat{\theta}(Y) - \Theta\| \mid \mathcal{L}\Big] \ge \frac{R}{2}\,\mathbb{P}(\widehat{J} \neq J \mid \mathcal{L}).$$

Let $P_j = N(Re_j, \tau^2 I_{M+N+1})$, let $\bar{P} = M^{-1}\sum_{j=1}^M P_j$, and take the reference law $Q = N(0, \tau^2 I_{M+N+1})$. Then

$$I(J;Y) = \frac{1}{M}\sum_{j=1}^M D_{\mathrm{KL}}(P_j\|\bar{P}) \le \frac{1}{M}\sum_{j=1}^M D_{\mathrm{KL}}(P_j\|Q) = \frac{R^2}{2\tau^2} = \frac{1}{2}\log M,$$

where the last equality uses $\tau = R/\sqrt{\log M}$. Fano's inequality therefore gives

$$\mathbb{P}(\widehat{J} \neq J \mid \mathcal{L}) \ge 1 - \frac{I(J;Y) + \log 2}{\log M} \ge c_0$$

for a universal constant $c_0 > 0$, for all sufficiently large $M$. Since $\mathbb{P}(\mathcal{L}) = 1/2$ under the original prior,

$$\mathbb{E}\|\widehat{\theta}(Y) - \Theta\| \ge \frac{1}{2}\,\mathbb{E}\Big[\|\widehat{\theta}(Y) - \Theta\| \mid \mathcal{L}\Big] \ge cR.$$

Taking the infimum over $\widehat{\theta}$ gives the Bayes-risk lower bound $cR$. Thus the class-wide Fano relaxation below is vacuous only as a certificate; it does not reflect an easy Bayes problem.

By allowing the hub mass to be placed at different points of $H_M = \{Re_0, Re_1, \dots, Re_M\}$, rather than only at the original hub $Re_0$, the same rotationally symmetric class construction can, for sufficiently large $M$, accommodate arbitrarily large finite families of priors that exhibit the separation and have markedly different local mass profiles. Consequently, the Bayes-optimal oracle may vary substantially from prior to prior, while the unrestricted estimator remains unchanged. This proves the statement about the Bayes-optimal risk in part (iii).

**Usual Fano relaxation is vacuous for the fixed prior.** Replacing the exact ghost mass by the worst-case relaxation $q_\pi(\Delta) = 1/2$ gives

$$\Delta\left[1 - \frac{\mathsf{I}_{\pi,Q_\tau} + \log 2}{\log(1/q_\pi(\Delta))}\right]_+ = \Delta\left[1 - \frac{\mathsf{I}_{\pi,Q_\tau} + \log 2}{\log 2}\right]_+ = 0.$$

Thus, for the same prior, channel, and estimator, the usual fixed-prior Fano relaxation is vacuous, while the exact algorithmic ghost mass gives a sharp Bayes estimation lower bound.

**Full-class minimax risk has arbitrarily large gap.** Let

$$\mathfrak{R}_{\mathrm{mm}}(T_{M,N},\tau) := \inf_{\widehat{\theta}} \sup_{\theta\in T_{M,N}} \mathbb{E}_\theta d(\theta,\widehat{\theta}(Y)),$$

where the infimum is over all measurable estimators with values in the ambient Euclidean space. The trivial estimator $\widehat{\theta}\equiv 0$ gives $\mathfrak{R}_{\mathrm{mm}}(T_{M,N},\tau)\le S$. Conversely, restrict the parameter to the cloud $C_N$ and put the uniform prior on $C_N$. For any estimator $\widehat{\theta}$, define

$$\widehat{J}(\widehat{\theta}) \in \arg\min_{1\le j\le N} \|\widehat{\theta}-Sf_j\|_2.$$

If $d(Sf_J,\widehat{\theta})<S/2$, then $\widehat{J}(\widehat{\theta})=J$, since the balls $B(Sf_j,S/2)$ are disjoint. Therefore

$$\mathbb{E}_J d(Sf_J,\widehat{\theta}(Y)) \ge \frac{S}{2}\,\mathbb{P}(\widehat{J}\ne J).$$

By Fano's inequality with reference law $N(0,\tau^2 I_{M+N+1})$,

$$I(J;Y)\le \frac{S^2}{2\tau^2}=\frac{1}{128}\log N.$$

Thus, for all sufficiently large $N$,

$$\mathbb{P}(\widehat{J}\ne J)\ge 1-\frac{I(J;Y)+\log 2}{\log N}\ge \frac{1}{2}.$$

It follows that

$$\mathfrak{R}_{\mathrm{mm}}(T_{M,N},\tau)\ge cS.$$

Together with the trivial upper bound, this gives $\mathfrak{R}_{\mathrm{mm}}(T_{M,N},\tau)\asymp S$, proving part (iv). Since $S/R=(1/8)\sqrt{\log N/\log M}$, the full-class minimax risk is much larger than the Bayes risk whenever $\log N/\log M\to\infty$. The minimax scale is governed by an ambient cloud that is irrelevant to the specified Bayesian decision problem.

**Algorithmically induced Gaussian certificate matches the selected subatlas.** Now define

$$X_t := \langle Z,t\rangle, \qquad t\in T_{M,N},$$

using the same standard Gaussian vector $Z$ as in the location experiment. We verify the comparison condition in Theorem 1.5. The decoder minimizes $\|Y-a\|^2+\lambda\|a\|^2$, so (1.14) holds with $\Omega(t)=\lambda\|t\|^2$ once $\Omega(\widehat{\Theta}_\lambda)\ge\Omega(\Theta)$ is checked. But $\Theta\in H_M$ almost surely, so $\|\Theta\|=R$, whereas every output of $\widehat{\Theta}_\lambda$ has norm either $R$ or $S\ge R$. Hence $\Omega(\widehat{\Theta}_\lambda)\ge\Omega(\Theta)$ almost surely. Applying Theorem 1.5 gives

$$\mathcal{G}_{\mathrm{alg}} := \mathbb{E}_{\Theta,Y}\left[X_{\widehat{\Theta}_\lambda(Y)}-X_\Theta\right]\gtrsim R\sqrt{\log M}.$$

Conversely, on the event $\widehat{\Theta}_\lambda\in H_M$, both $\Theta$ and $\widehat{\Theta}_\lambda$ lie in $H_M$, and

$$\mathbb{E}\sup_{s,u\in H_M}(X_s-X_u)\asymp R\sqrt{\log M}.$$

The event $\widehat{\Theta}_\lambda\in C_N$ has probability at most $N^{-2}$ under the true prior. By Cauchy–Schwarz and the standard Gaussian maximal bound on $T_{M,N}$, its contribution is negligible. Therefore

$$\mathcal{G}_{\mathrm{alg}}\asymp \mathbb{E}\sup_{s,u\in H_M}(X_s-X_u)\asymp R\sqrt{\log M},$$

which proves the first assertion in part (v).

This also matches the pointwise Gaussian upper envelope on the selected subatlas. By (4.1), we already have

$$Q_\tau\{\widehat{\Theta}_\lambda(Y) \in C_N\} \le N^{-2}.$$

for all sufficiently large $N$. Let

$$\nu_Q^H(a) := Q_\tau\{\widehat{\Theta}_\lambda(Y) = a \mid \widehat{\Theta}_\lambda(Y) \in H_M\}, \qquad a \in H_M,$$

be the ghost selection law conditioned on the lower-norm subatlas. By exchangeability, $\nu_Q^H$ is uniform on $H_M$. Since the restricted process on $H_M$ consists of $M+1$ Gaussian variables of variance at most $R^2$, Theorem 1.1 gives, up to universal constants and confidence logarithms,

$$\mathbb{P}\left(\forall t \in H_M : \ |X_t| \le CR\sqrt{\log \frac{M}{\delta}}\right) \ge 1-\delta.$$

This proves the pointwise-envelope assertion in part (v).

**Global Gaussian supremum has arbitrarily large gap.** The full ambient Gaussian field has a larger global scale because of the cloud. Comparing the cloud points with the fixed hub $Re_0$,

$$\mathbb{E} \sup_{s,u\in T_{M,N}} (X_s - X_u) \ge \mathbb{E}\left[\max_{1\le j\le N} X_{Sf_j} - X_{Re_0}\right] = S\,\mathbb{E} \max_{1\le j\le N} Z_j \gtrsim S\sqrt{\log N}.$$

Using the definition of $S$,

$$S\sqrt{\log N} = \frac{1}{8} R \frac{\log N}{\sqrt{\log M}}.$$

This proves part (vi). Comparing with the decision-aligned scale $R\sqrt{\log M}$, the gap is of order $\log N/\log M$, and taking $N = M^K$ makes the ratio of order $K$.

□

**Summary and interpretation.** The construction gives a fully explicit sequence of finite-dimensional Gaussian location problems for which

$$\begin{aligned} &\text{fixed-prior Fano relaxation} \\ \ll\ &\text{algorithmic ghost-mass lower envelope} \\ \asymp\ &\text{pointwise upper envelope on the selected subatlas} \\ \ll\ &\text{full-class minimax/global Gaussian scale.} \end{aligned}$$

The example does not claim that minimax theory is wrong or fundamentally insufficient. Rather, it shows that a full-class minimax benchmark can be the wrong certificate for a pointwise decision problem, while a restricted minimax benchmark can be oracle-dependent. The norm-regularized decoder uses a fixed numerical penalty $\lambda$, independent of the prior $\pi$ and the problem parameters $M, N, R, S, \tau$, while identifying the hub–leaf subatlas $H_M$ as the relevant subatlas is precisely the decision-dependent information revealed by the ghost image $Y \mapsto \widehat{\Theta}_\lambda(Y)$. The algorithmic ghost-mass lower bound supplies the missing certificate: it is tied to the fixed prior, the actual estimator, and the subgeometry selected by that estimator. This is the finite-dimensional analogue of the pointwise-complexity viewpoint in deep networks, where the effective subatlas is induced by the learned representation and compressed spectrum rather than specified as a fixed restricted class in advance.

Here we keep the cleanest finite formulation. Finite sets are already the standard setting for Sudakov, Fano, and packing lower bounds, and they make the quantifiers transparent: the full ambient class, the oracle restricted subatlas, the fixed prior, and the actual estimator are all explicit. Extensions to continuous settings, as well as computational-efficiency questions concerning the full class versus selected subatlases, are left to future work.

# 5 Conclusion

The paper develops three linked pointwise and algorithmic messages, and shows how they yield separations from global complexity scales. First, for Gaussian processes, the natural refinement of generic chaining is a simultaneous pointwise envelope for the field, not only a scalar estimate on $\mathbb{E}\sup_x X_x$. This envelope is sharp in expectation after optimizing over the prior and recovers the anchored Fernique–Talagrand scale. It should be useful whenever one wants to retain local field information before taking a final supremum.

Second, the Bayesian algorithmic lower envelope gives a single-radius lower-bound counterpart to pointwise complexity. It is an information-theoretic Bayes-risk bound for every estimator $\widehat{t}(Y)$, expressed through the exact ghost small-ball mass $\mathbb{E}_Q \pi(B_d(\widehat{t}(Y), \Delta))$. The comparison-decoder theorem turns this Bayes distance obstruction into a lower bound on a decision-aligned Gaussian range; conversely, a simultaneous pointwise Gaussian envelope yields an algorithmic upper bound on the same decoding risk. Importantly, we construct a weighted-basis "hub–leaves–cloud" example for norm-regularized nearest-neighbor decoder where the fixed-prior Fano relaxation is vacuous, the algorithmic ghost-mass lower envelope and pointwise upper envelope match on the estimator-selected subatlas, while the full-class minimax risk and global Gaussian scale are much larger. The key is to retain the algorithmic mass: it provides the local certificate of pointwise complexity for fixed estimators in overparameterized ambient classes, precisely in regimes where classical minimax theory becomes either too coarse or oracle-dependent.

Third, finite-cutoff renormalization and graph local time illustrate the same structural mechanism in settings motivated by physics and probability. The structural analogies between RG and DNN, and between pointwise dimension and graph local time, show that trajectory-aligned or selected-subatlas complexity can be intrinsically smaller than the scale obtained by passing through a full-cover event.

Taken together, the results suggest a unified principle appears: global worst-case objects capture genuine hardness of the full class, while pointwise and algorithmic objects can certify tractability on the local geometry selected by a field, an estimator, or a renormalization trajectory. In this sense, the paper connects field-level generic chaining to decision- and trajectory-aligned certificates, thereby establishing separations between pointwise and global complexity scales.

# A Finite-Cutoff Renormalization Application

## A.1 Structural analogy between finite-cutoff RG and feature-learning DNN

This subsection makes precise the analogy between finite-cutoff renormalization and feature-learning DNN in Li and Xu (2026), while also recording where the analogy stops. The analogy does not rest merely on the correspondence between composite layers and RG levels. More essentially, both settings exhibit a common *feature/operator inner-product* structure within each composite layer of a DNN, or within each exponential-family RG level. This correspondence leads to the same proof mechanism: exact finite-step telescoping, a pointwise ellipsoidal metric induced by a learned or renormalized secant Gram, and a hierarchical atlas that separates local effective dimension from the global price of making the prior independent of the realized trajectory.

The distinction is also important. A DNN layer has a rectangular matrix parameter and a common input-feature *matrix*, so its pointwise metric has a Kronecker repetition. A Wilsonian RG step is the operation of integrating out short-distance or fluctuation variables and rewriting the resulting coarse marginal as an effective action. After a finite operator truncation, this effective action is usually represented by a *vector* of couplings $g_j = (g_{j,1}, \ldots, g_{j,p_j})$, not by a learned matrix weight. An exponential-family RG coordinate system is one standard finite-dimensional way to write this truncation: one chooses operators $O_{j,a}$ and lets the couplings $g_{j,a}$ be natural parameters. It is therefore best understood as a coordinate/projection model for a finite Wilsonian step, not as a new claim that RG has neural-network-style matrices. We clarify this distinction after (A.4) and again after Theorem A.4.

The practical value in statistical field theory is that RG is fundamentally about controlling coarse marginals after integrating out fluctuations. For globally non-log-concave and anisotropic interacting fields, direct scalar Lipschitz bounds or global convex-geometric arguments are often too crude. The finite-cutoff viewpoint instead allows one to work on stable local charts where the effective action has a curvature lower bound: directions below the finite resolution are absorbed as approximation error, while global nonconvexity is paid for through an atlas cost. We formalize this mechanism as an abstract deterministic complexity theorem and illustrate it through a chartwise nonconvex phase-atlas example, as discussed at the end of Subsection A.3.

| Deep-network pointwise theory | Finite-cutoff RG analogue |
|---|---|
| Layer index $\ell$ | RG level index $j$ |
| Weights matrix $W_\ell$ | Level coupling vector $g_j$ |
| Feature matrix $F_{\ell-1}(W, X)$ | Renormalized operator family $O_j(S_j)$ |
| Feature Gram $F_{\ell-1}F_{\ell-1}^\top$ | Secant response Gram $\Gamma_j^{\mathrm{sec}}$ |
| Exact layer telescoping | Exact replacement over RG levels |
| Outer Lipschitz $M_{\ell\to L}$ | Stability of later RG maps $M_{j\to J}$ |
| Active feature eigenspace | Relevant/marginal active operator subspace |
| Grassmannian prior | Phase/subspace atlas prior |
| Feature-rank compression | Contraction or truncation of irrelevant directions |
| Kernel/NTK baseline | Free or massive GFF baseline |

**Terminology for radii, resolutions, and RG levels.** The word "scale" appears in several neighboring literatures, so we keep the terminology separate. The Bayesian lower envelope is a *single-radius* statement at a testing radius $\Delta$, whereas generic chaining is multiresolution and integrates over radii. The Gaussian upper theorem also uses neighborhood radii, such as $\varepsilon$, $r_0$, and $s_0$, only as metric resolutions. In the renormalization appendix we use *RG level* for the index of a block-spin or fluctuation integration, and *finite-cutoff* or *finite-step* for the non-infinitesimal nature of the telescoping argument. Finally, in the separating example the numbers $R$ and $S$ are norms of two different parts of the ambient class. These are related by analogy but are not the same mathematical object.

**Limiting scope relative to constructive QFT, RG approximation, and gauge fields.** The language is compatible with rigorous Wilsonian RG frameworks such as Brydges and Slade (2015); Bauerschmidt et al. (2019), but all statements are abstract finite-dimensional and finite-cutoff statements: they provide a structural complexity layer, not a replacement for model-specific constructive-QFT estimates. In the four-dimensional Ising and lattice $\Phi_4^4$ setting, Aizenman and Duminil-Copin (2021) prove marginal triviality by controlling deviations from Wick's rule through random currents and multiscale estimates; the issue is not amplitude control, since Ising spins are already uniformly bounded, but Gaussianization of macroscopic observables through decay of connected non-Gaussian remainders. Regarding the technical difficulty of Gaussianization, while universal approximation theorems (Cybenko, 1989; Hornik et al., 1989) can approximate a finite-dimensional RG map on a compact set at fixed cutoff, they do not provide scale-uniform error bounds, preservation of locality or reflection positivity, or control of Wick-factorization errors. For Yang–Mills, an analogous finite-cutoff program would have to be formulated for lattice gauge fields and gauge-invariant observables such as plaquettes, Wilson loops, and character observables; a mass-gap theorem would still require continuum construction, Osterwalder–Schrader positivity and reconstruction, and uniform exponential clustering of gauge-invariant correlations (Jaffe and Witten, 2000; Chatterjee, 2018). Thus scaling-limit Gaussianity, nontrivial continuum interaction, and gauge-theoretic mass gaps remain separate model-specific problems.

### A.2 Pointwise complexity of composite renormalization maps

**Exact finite-step telescoping for composite renormalization maps.** Let $\mathcal{S}_0, \dots, \mathcal{S}_J$ be finite-dimensional normed spaces. For each RG level $j$, let

$$\mathcal{R}_j^{g_j} : \mathcal{S}_j \to \mathcal{S}_{j+1}, \qquad g_j \in B_2(R_j) \subset \mathbb{R}^{p_j},$$

be a nonlinear coarse-graining map. Starting from a fixed $S_0 \in \mathcal{S}_0$, define the RG trajectory

$$S_{j+1}(g) := \mathcal{R}_j^{g_j}(S_j(g)), \qquad S_J(g) = \mathcal{R}_{J-1}^{g_{J-1}} \circ \cdots \circ \mathcal{R}_0^{g_0}(S_0).$$

This notation covers, for example, the exact Wilsonian identity

$$e^{-S_{j+1}(\Phi)} = \int e^{-S_j(\Phi+\psi_j)}\, d\gamma_j(\psi_j).$$

This is the finite-dimensional analogue of integrating out one fluctuation field at one RG level. Perturbative RG expands this map; here we keep the finite map itself.

This is the sense in which the appendix uses Wilsonian RG (Brydges and Slade, 2015; Bauerschmidt et al., 2019) . Wilsonian RG is the coarse-graining map on actions or measures obtained by integrating out fluctuations. Exponential-family RG is the finite-coordinate representation obtained after choosing an operator family $O_{j,a}$ and writing the retained effective action in natural parameters $g_{j,a}$. The exact Wilsonian step may generate infinitely many operators; the exponential-family description is exact only if the chosen family is closed under the step, and otherwise it is a finite-dimensional projection or truncation. In both cases the ordinary finite coordinate is a vector of couplings.

For two coupling sequences $g, g'$, define the hybrid trajectory

$$S_j^{(g',g;m)} := \mathcal{R}_{j-1}^{g'_{j-1}} \circ \cdots \circ \mathcal{R}_m^{g'_m} \circ \mathcal{R}_{m-1}^{g_{m-1}} \circ \cdots \circ \mathcal{R}_0^{g_0}(S_0),$$

with the evident convention at the endpoints. Then the following identity is exact.

**Lemma A.1** (Exact RG replacement telescoping)**.** *For every $g, g'$,*

$$S_J(g') - S_J(g) = \sum_{j=0}^{J-1} \Big[ \mathcal{R}^{g'}_{J-1:j+1}\big(\mathcal{R}_j^{g'_j}(S_j^{(j)})\big) - \mathcal{R}^{g'}_{J-1:j+1}\big(\mathcal{R}_j^{g_j}(S_j^{(j)})\big)\Big], \tag{A.1}$$

*where $S_j^{(j)} := \mathcal{R}_{j-1}^{g_{j-1}} \circ \cdots \circ \mathcal{R}_0^{g_0}(S_0)$ and $\mathcal{R}^{g'}_{J-1:j+1} := \mathcal{R}_{J-1}^{g'_{J-1}} \circ \cdots \circ \mathcal{R}_{j+1}^{g'_{j+1}}$.*

*Proof.* Insert the hybrid trajectories that use $g$ up to level $j-1$ and $g'$ from level $j$ onward. The difference between two consecutive hybrids changes only the scale-$j$ map. Summing these consecutive differences gives (A.1). ☐

Let $\rho$ be a pseudometric (in particular, satisfying the triangle inequality) on $\mathcal{S}_J$. The next result is the RG analogue of the non-perturbative DNN feature expansion in Li and Xu (2026, Lemma 1): the finite replacement identity induces an ellipsoidal metric whose matrices are finite-scale secant response Grams.

**Theorem A.2** (Secant-Gram metric domination for nonlinear RG)**.** *Fix $g$ and a finite neighborhood $\mathcal{N}(g,\varepsilon)$. Assume that for each $j$:*

*(i) the later RG flow is locally stable: for every $g' \in \mathcal{N}(g,\varepsilon)$ and all states $U, V$ arising along the corresponding hybrid trajectories,*

$$\rho\big(\mathcal{R}^{g'}_{J-1:j+1}(U), \mathcal{R}^{g'}_{J-1:j+1}(V)\big) \le M_{j\to J}(g,\varepsilon)\|U-V\|_{j+1};$$

*(ii) the scale-$j$ finite replacement is dominated by a positive semidefinite secant Gram $\Gamma_j(g,\varepsilon)$:*

$$\big\|\mathcal{R}_j^{g'_j}(S_j^{(j)}) - \mathcal{R}_j^{g_j}(S_j^{(j)})\big\|_{j+1}^2 \le (g'_j - g_j)^\top \Gamma_j(g,\varepsilon)(g'_j - g_j).$$

*Then all $g' \in \mathcal{N}(g,\varepsilon)$ satisfy*

$$\rho(S_J(g'), S_J(g))^2 \le \sum_{j=0}^{J-1} (g'_j - g_j)^\top G_j(g,\varepsilon)(g'_j - g_j), \qquad G_j := J\, M_{j\to J}(g,\varepsilon)^2 \Gamma_j(g,\varepsilon). \tag{A.2}$$

*Proof.* Apply Lemma A.1, the triangle inequality for $\rho$, the local stability assumption, and the one-step secant-Gram bound. Cauchy–Schwarz gives a factor $J$ when the $J$ increments are summed; this factor is included in the definition of $G_j$. ☐

**How the secant Gram is verified in exponential-family RG steps.** Condition (ii) in Theorem A.2 is a finite-dimensional secant calculation. Fix the state $S_j^{(j)}$ and set

$$F_j(g_j) := \mathcal{R}_j^{g_j}(S_j^{(j)}).$$

If $F_j$ is continuously Frechet differentiable on the finite neighborhood under consideration and the next-scale norm is Hilbertian, then for $\Delta_j = g'_j - g_j$ and $g_{j,t} = g_j + t\Delta_j$,

$$F_j(g'_j) - F_j(g_j) = \int_0^1 DF_j(g_{j,t})[\Delta_j]dt,$$

and Jensen's inequality gives

$$\|F_j(g'_j) - F_j(g_j)\|_{j+1}^2 \le \Delta_j^\top \left(\int_0^1 DF_j(g_{j,t})^* DF_j(g_{j,t})dt\right)\Delta_j. \tag{A.3}$$

Thus condition (ii) holds with the pairwise secant Gram in parentheses, or with any positive-semidefinite Loewner envelope that dominates it uniformly over the finite neighborhood. For an exponential-family RG step

$$\mathcal{R}_j^g(S)(\Phi) = -\log \int \exp\Big\{-S(\Phi+\psi) - \sum_{a=1}^{p_j} g_a O_{j,a}(\Phi+\psi)\Big\} d\gamma_j(\psi),$$

one has

$$\partial_{g_a}\mathcal{R}_j^g(S)(\Phi) = \mathbb{E}_{j,\Phi,g}[O_{j,a}(\Phi+\psi)].$$

Consequently the secant Gram is the Gram matrix of the finite-scale response functions

$$\Gamma^{\mathrm{sec}}_{j,ab}(g,g') = \int_0^1 \langle m^t_{j,a}, m^t_{j,b}\rangle_{j+1} dt, \qquad m^t_{j,a}(\Phi) := \mathbb{E}_{j,\Phi,g_{j,t}}[O_{j,a}(\Phi+\psi)]. \tag{A.4}$$

This is the precise RG analogue of the learned feature Gram matrix in the DNN pointwise metric. The word "analogue" here means "response-Gram analogue", not "identical matrix-layer analogue". The DNN metric block is $F_{\ell-1}F_{\ell-1}^\top \otimes I_{d_\ell}$ because the same incoming feature Gram is repeated over $d_\ell$ output rows. The generic RG secant Gram $\Gamma_j^{\mathrm{sec}} \in \mathbb{R}^{p_j \times p_j}$ acts on the coupling vector $g_j$. Without additional channel structure it does not produce the Kronecker repetition that balances DNN local and atlas costs.

**Free Gaussian RG as the fixed-kernel baseline** Let $E = E_C \oplus E_F$ be finite-dimensional and let a centered Gaussian field have precision matrix

$$Q = \begin{pmatrix} Q_{CC} & Q_{CF} \\ Q_{FC} & Q_{FF} \end{pmatrix}, \qquad Q_{FF} \succ 0.$$

Define the Schur-complement precision

$$Q_C^{\mathrm{RG}} := Q_{CC} - Q_{CF}Q_{FF}^{-1}Q_{FC}.$$

Then marginalizing the fluctuation variables gives $\phi_C \sim N(0, (Q_C^{\mathrm{RG}})^{-1})$, and conditionally

$$\phi_F = -Q_{FF}^{-1}Q_{FC}\phi_C + \zeta, \qquad \zeta \sim N(0, Q_{FF}^{-1})$$

independently of $\phi_C$. This is a fixed-kernel RG step: all pointwise complexity is determined by the Schur-complement covariance and the fluctuation covariance. It is therefore analogous to a kernel, GP, or NTK model rather than to nonlinear feature learning.

**Hierarchical phase/subspace prior.** For a positive semidefinite matrix $G \in \mathbb{R}^{p\times p}$, a Euclidean radius $R$, and a resolution $u > 0$, set the effective rank and the effective subspace

$$r_{\text{eff}}(G,R,u) := \max\{k : \lambda_k(G)R^2 \geq u^2\}, \qquad V_{\text{eff}}(G,R,u) := \text{span}\{\text{top } r_{\text{eff}} \text{ eigenvectors}\}$$

with the convention $\max\emptyset := 0$. Define the effective dimension

$$d_{\text{eff}}(G,R,u) := \frac{1}{2}\sum_{k=1}^{r_{\text{eff}}(G,R,u)} \log\left(\frac{8R^2\lambda_k(G)}{u^2}\right). \tag{A.5}$$

The following elementary estimate is the local-chart calculation used throughout. For subspaces $V, \bar{V} \subset \mathbb{R}^p$, write

$$\rho_{\text{proj},G}(V,\bar{V}) := \|G^{1/2}(P_V - P_{\bar{V}})\|_{\text{op}}.$$

**Lemma A.3** (Ellipsoidal local chart)**.** *Let $r = r_{\text{eff}}(G,R,u)$, let $V = V_{\text{eff}}(G,R,u)$, and let $\bar{V} \in \text{Gr}(p,r)$ satisfy*

$$\rho_{\text{proj},G}(V,\bar{V}) \leq \frac{u}{4R}.$$

*Let $\pi_{\bar{V}}$ be the uniform law on $B_2(2R)\cap\bar{V}$. Then uniformly over $g \in B_2(R)$,*

$$\log\frac{1}{\pi_{\bar{V}}\{g' : (g'-g)^\top G(g'-g) \leq u^2\}} \leq \frac{1}{2}\sum_{k=1}^{r_{\text{eff}}(G,R,u)} \log\left(\frac{40R^2\lambda_k(G)}{u^2}\right) \leq 1.5 d_{\text{eff}}(G,R,u).$$

*Equivalently, replacing $u$ by a constant multiple gives the displayed local-chart bound used in Theorem A.4.*

Lemma A.3 is based on the same finite-dimensional volume-ratio argument used in the DNN local-chart bound of Li and Xu (2026, Lemma 2); we refer to Li and Xu (2026, Section C.3) for the proof. There are only two minor changes. First, for simplicity, we enlarge the global radius of the prior support from 1.58R to 2R, which does not affect the argument. Second, in the final step we use the cutoff definition (A.5) of the effective dimension to obtain

$$\frac{1}{2}\sum_{k=1}^{r_{\text{eff}}(G,R,u)} \log\left(\frac{40R^2\lambda_k(G)}{u^2}\right) = d_{\text{eff}}(G,R,u) + \frac{\log 5}{2}\, r_{\text{eff}}(G,R,u) \leq 1.5\, d_{\text{eff}}(G,R,u).$$

This observation also shows that the absolute-constant rescaling of the radius in effective-dimension terms of the form $d_{\text{eff}}(G, C\cdot R, u)$ in Li and Xu (2026) can be refined to an absolute-constant multiplicative factor $C \cdot d_{\text{eff}}(G,R,u)$, while keeping the effective-rank term at the original radius R. We leave the systematic incorporation of this refinement into the formal statements of Li and Xu (2026) and related follow-up work to future versions.

Let $\mathcal{A}_j$ be a finite set of phase labels at level $j$, with prior weights $q_j(a) > 0$. Conditional on a phase label $a$ and rank $r$, let $\nu_{j,a,r}$ be a prior on the Grassmannian $\text{Gr}(p_j, r)$. For $G_j = G_j(g,\varepsilon)$, define the atlas cost

$$\mathfrak{A}_j(g,\varepsilon,u_j) := \log\frac{1}{q_j(a_j(g))} + \log(1+p_j) + \log\frac{1}{\nu_{j,a_j,r_j}\left(B_{\rho_{\text{proj},G_j}}(V_j, u_j/(4R_j))\right)}, \tag{A.6}$$

where $r_j = r_{\text{eff}}(G_j,R_j,u_j)$, $V_j = V_{\text{eff}}(G_j,R_j,u_j)$, and $a_j(g)$ is any chart containing the trajectory at level $j$.

**Theorem A.4** (Finite-scale RG pointwise dimension). *Assume the metric domination (A.2). Choose resolutions $u_j > 0$ such that $\sum_{j=0}^{J-1} u_j^2 \le c\varepsilon^2$. Construct a trajectory-independent hierarchical prior $\Pi$ by independently, at each level $j$, sampling a phase label $a$, an effective rank $r$, a reference subspace $\bar{V} \in \mathrm{Gr}(p_j, r)$, and then sampling $g_j$ uniformly in $B_2(2R_j) \cap \bar{V}$. Then, there exists an absolute constant $C > 0$ such that uniformly for every trajectory $g$,*

$$\log \frac{1}{\Pi(B_\rho(g,\varepsilon))} \le C \sum_{j=0}^{J-1} \left[ d_{\mathrm{eff}}(G_j(g,\varepsilon), R_j, u_j) + \mathfrak{A}_j(g,\varepsilon,u_j) \right]. \tag{A.7}$$

*Proof.* For each scale choose the phase label $a_j(g)$, the active rank $r_j$, and a reference subspace $\bar{V}_j$ within projection distance $u_j/(4R_j)$ of $V_j$. Lemma A.3 gives the local prior mass of the ellipsoidal $u_j$-ball inside this chart. Multiplying the scale-wise prior masses and incorporating the phase, rank, and subspace probabilities yields the right-hand side of (A.7). For a detailed explanation of the corresponding decomposition, we refer to Li and Xu (2026, Section 3.3.3). Finally, if each scale perturbation lies in its ellipsoidal $u_j$-ball, then (A.2) and $\sum_j u_j^2 \le c\varepsilon^2$ imply $g' \in B_\rho(g,\varepsilon)$. Hence the product event is contained in the $\rho$-ball, and its prior mass lower-bounds $\Pi(B_\rho(g,\varepsilon))$. □

The theorem is the RG counterpart of the DNN Riemannian-dimension calculation. The local term $d_{\mathrm{eff}}$ is the effective dimension of the finite-scale response Gram. The atlas term is the price of making the prior independent of the realized RG trajectory. Irrelevant directions do not contribute once their eigenvalues fall below the finite resolution; relevant and marginal directions are precisely the active directions that remain in the sum.

The theorem deliberately leaves $\mathfrak{A}_j$ explicit. As in the DNN setting, structural mechanisms are needed to make the global-atlas cost comparable to the local-chart cost, but the mechanism is model-dependent. In DNN, this balance comes from the matrix-weight/common-feature structure and the resulting Kronecker-repetition metric. In genuine RG settings, it must instead come from physical structure in the operator atlas, with the nonconvex example below providing a clean instance of cost balancing without a DNN-style matrix structure.

### A.3 A finite nonconvex phase-atlas example

The previous theorem is abstract. We now give a concrete finite-volume chartwise example. The point is not to construct a full continuum double-well field, but to show how a globally non-log-concave finite measure can be controlled after it is decomposed into stable phase charts. Let $E = E_C \oplus E_F$, let $\mathcal{A}$ be a finite set of phases, and let

$$\nu(d\phi_C, d\phi_F) = \sum_{a \in \mathcal{A}} w_a \nu_a(d\phi_C, d\phi_F), \qquad \nu_a \propto e^{-S_a(\phi_C,\phi_F)} \mathbf{1}_{\Omega_{C,a} \times \Omega_{F,a}} d\phi_C d\phi_F,$$

where the sets $\Omega_{C,a}, \Omega_{F,a}$ are closed convex sets with nonempty interior. The mixture may be globally non-log-concave and multimodal. Assume that, on each chart,

$$\nabla^2 S_a(\phi_C,\phi_F) \succeq H_a = \begin{pmatrix} A_a & B_a \\ B_a^\top & D_a \end{pmatrix}, \qquad D_a \succ 0, \qquad H_a^{\mathrm{RG}} := A_a - B_a D_a^{-1} B_a^\top \succ 0.$$

**Theorem A.5** (One-step nonconvex phase-atlas RG stability). *Under the preceding assumptions, the coarse marginal is again a finite phase-atlas mixture*

$$\nu_C(d\phi_C) = \sum_{a \in \mathcal{A}} w_a^+ \nu_{C,a}(d\phi_C), \qquad \nu_{C,a} \propto e^{-S_a^{\mathrm{RG}}(u)} \mathbf{1}_{\Omega_{C,a}} du,$$

*where the weights* $w_a^+$ *are the corresponding normalized marginal weights (in the convention that each* $\nu_a$ *is already normalized,* $w_a^+ = w_a$*) and*

$$S_a^{\mathrm{RG}}(u) := -\log \int_{\Omega_{F,a}} e^{-S_a(u,v)} dv.$$

*Moreover,* $S_a^{\mathrm{RG}}$ *is* $H_a^{\mathrm{RG}}$*-strongly convex. Hence, for any finite family of coarse linear observables* $Y_x(u) = \langle v_x, u\rangle$*, the centered chart process* $Y_x - \mathbb{E}_{\nu_{C,a}} Y_x$ *is sub-Gaussian with canonical metric*

$$d_a(x,y)^2 = \langle v_x - v_y, (H_a^{\mathrm{RG}})^{-1}(v_x - v_y)\rangle.$$

*Consequently, the standard sub-Gaussian chaining and variance-peeling arguments, applied chart by chart with the comparison metric induced by* $(H_a^{\mathrm{RG}})^{-1}$*, yield simultaneous chartwise envelopes. If one wants a single event valid on all phase charts, allocate confidence levels* $\delta_a$ *across charts; this adds a chart cost* $\log(1/\delta_a)$*, for instance* $\log(1/w_a^+)$ *under weight-proportional allocation or* $\log|\mathcal{A}|$ *under uniform allocation.*

*Proof.* The marginal formula is the definition of integration over $E_F$. The Schur-complement lower bound follows by completing the square:

$$\begin{pmatrix} u \\ v \end{pmatrix}^{\top} H_a \begin{pmatrix} u \\ v \end{pmatrix} = u^\top H_a^{\mathrm{RG}} u + (v + D_a^{-1}B_a^\top u)^\top D_a (v + D_a^{-1}B_a^\top u).$$

Moreover,

$$H_a - \begin{pmatrix} H_a^{\mathrm{RG}} & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} B_a D_a^{-1} B_a^\top & B_a \\ B_a^\top & D_a \end{pmatrix} \succeq 0.$$

Hence $S_a(u,v) - \frac{1}{2}u^\top H_a^{\mathrm{RG}} u$ is jointly convex on the product chart. Including the indicators of the convex sets $\Omega_{C,a}$ and $\Omega_{F,a}$, the Prékopa–Leindler theorem implies that $S_a^{\mathrm{RG}}(u) - \frac{1}{2}u^\top H_a^{\mathrm{RG}} u$ is convex on $\Omega_{C,a}$. Thus the coarse chart is $H_a^{\mathrm{RG}}$-strongly log-concave. Brascamp–Lieb and Herbst then give the stated sub-Gaussian bound for linear observables (Brascamp and Lieb, 1976). The envelope follows from the usual generic-chaining upper bound for sub-Gaussian increments, followed by the same pointwise peeling used in the Gaussian proof on each chart and the confidence allocation over $a \in \mathcal{A}$ described in the theorem. □

**Proposition A.6** (Phase-atlas bound versus global Lipschitz/full-cover bound)**.** *In the setting of Theorem A.5, suppose that the coarse observables are indexed by a finite full class* $\mathcal{T}_{\mathrm{full}}$*,*

$$Y_x(u) = \langle v_x, u\rangle, \qquad x \in \mathcal{T}_{\mathrm{full}},$$

*and that each phase* $a \in \mathcal{A}$ *comes with a selected subfamily* $\mathcal{T}_a \subseteq \mathcal{T}_{\mathrm{full}}$*. Write*

$$Z_{a,x} := Y_x - \mathbb{E}_{\nu_{C,a}} Y_x, \qquad \ell_a(x)^2 := \langle v_x, (H_a^{\mathrm{RG}})^{-1} v_x\rangle,$$

$$L_a := \sup_{x\in\mathcal{T}_a} \ell_a(x), \qquad L_{\mathrm{glob}} := \sup_{a\in\mathcal{A}} \sup_{x\in\mathcal{T}_{\mathrm{full}}} \ell_a(x), \qquad M_{\mathrm{full}} := |\mathcal{T}_{\mathrm{full}}|.$$

*Equivalently, if* $H_a^{\mathrm{RG}} \succeq \lambda_{\mathrm{glob}} I$ *for all* $a$ *and* $\|v_x\|_2 \le L_{\mathrm{Euc}}$*, then* $L_{\mathrm{glob}} \le L_{\mathrm{Euc}}/\sqrt{\lambda_{\mathrm{glob}}}$*, which is the usual global Lipschitz constant after the worst curvature normalization. Let* $q \in \Delta(\mathcal{A})$ *be an atlas prior with* $q(a) > 0$*. Then for every* $a \in \mathcal{A}$ *and* $\delta \in (0,1)$*,*

$$\nu_{C,a}\left\{ \sup_{x\in\mathcal{T}_a} |Z_{a,x}| \le C L_a \sqrt{\log \frac{2|\mathcal{T}_a|}{\delta q(a)}} \right\} \ge 1 - \delta q(a). \tag{A.8}$$

*By contrast, the global Lipschitz bound obtained by ignoring the phase/subatlas structure and covering the full class by singletons gives, uniformly for every* $a \in \mathcal{A}$*,*

$$\nu_{C,a}\left\{\sup_{x\in\mathcal{T}_{\mathrm{full}}} |Z_{a,x}| \le CL_{\mathrm{glob}}\sqrt{\log\frac{2M_{\mathrm{full}}}{\delta}}\right\} \ge 1-\delta. \tag{A.9}$$

*Thus the selected phase-atlas certificate pays the local class size and the code length* $\log(1/q(a))$*, while the global full-cover certificate pays the full class size* $M_{\mathrm{full}}$ *and the worst Lipschitz constant* $L_{\mathrm{glob}}$*. If the selected phase is fixed in advance and no trajectory-independent atlas code is needed, the factor* $q(a)$ *in* (A.8) *can be omitted. For infinite or continuous observable classes, the singleton cardinalities in* (A.8)–(A.9) *are replaced by the corresponding metric covering or chaining integrals; the same comparison is between a selected chart cover and a full class-wide cover.*

*Moreover, the separation can be strict. Fix integers* $N, K \ge 2$*. Let* $\mathcal{A} = \{1, \ldots, N\}$*, let* $E_C = \mathbb{R}^{N(K+1)}$ *with orthonormal coordinates* $e_{a,0}, e_{a,1}, \ldots, e_{a,K}$*, and put*

$$S_a^{\mathrm{RG}}(u) = \frac{1}{2}\|u - m_a\|_2^2, \qquad m_a = Be_{a,0}, \qquad B \ge 4\sqrt{\log N}.$$

*Then* $\nu_{C,a} = \mathcal{N}(m_a, I)$*,* $H_a^{\mathrm{RG}} = I$*, and, with weights* $w_1 = 1/2$ *and* $w_a = 1/[2(N-1)]$ *for* $a \ge 2$*, the mixture* $\sum_a w_a \nu_{C,a}$ *is globally non-log-concave. Define*

$$\mathcal{T}_{\mathrm{full}} := \{(b,k) : 1 \le b \le N, 1 \le k \le K\}, \qquad Y_{b,k}(u) = \langle e_{b,k}, u\rangle, \qquad \mathcal{T}_a := \{(a,k) : 1 \le k \le K\}.$$

*For the selected phase* $a_\star = 1$*, choose an atlas prior with* $q(1) \ge q_0 > 0$*. Then* $L_{a_\star} = L_{\mathrm{glob}} = 1$*,*

$$\mathbb{E}_{\nu_{C,1}} \sup_{x\in\mathcal{T}_1} |Z_{1,x}| \asymp \sqrt{\log K}, \tag{A.10}$$

*whereas*

$$\mathbb{E}_{\nu_{C,1}} \sup_{x\in\mathcal{T}_{\mathrm{full}}} |Z_{1,x}| \asymp \sqrt{\log(NK)}. \tag{A.11}$$

*Consequently, if* $\log N/\log K \to \infty$ *and* $q(1)$ *stays bounded below, the global Lipschitz/full-cover scale is larger than the selected phase-atlas scale by an unbounded factor. If instead the atlas prior is uniform over all phases, then* $\log(1/q(a)) = \log N$*, and this particular atlas advantage disappears; the separation is precisely a statement about a nonuniform or trajectory-aligned phase code, not about a free reduction of an arbitrary unknown phase.*

*Proof.* By the strong log-concavity conclusion of Theorem A.5, the Herbst argument gives, for every $a$ and $x$,

$$\mathbb{E}_{\nu_{C,a}} \exp\{\lambda Z_{a,x}\} \le \exp\left(\frac{\lambda^2 \ell_a(x)^2}{2}\right), \qquad \lambda \in \mathbb{R}.$$

Hence

$$\nu_{C,a}\{|Z_{a,x}| > t\} \le 2\exp\left(-\frac{t^2}{2\ell_a(x)^2}\right),$$

with the evident deterministic interpretation when $\ell_a(x) = 0$. A union bound over $x \in \mathcal{T}_a$, followed by the choices

$$t = CL_a\sqrt{\log\frac{2|\mathcal{T}_a|}{\delta q(a)}} \quad \text{and} \quad t = CL_{\mathrm{glob}}\sqrt{\log\frac{2M_{\mathrm{full}}}{\delta}},$$

proves (A.8) and (A.9), after adjusting the universal constant $C$. This is the full-cover proof: it uses only the worst Lipschitz scale and the number of points in the full observable class, and does not exploit the selected subfamily $\mathcal{T}_a$.

It remains to verify the strict example. The Hessian of each quadratic chart is exactly the identity, so all chart assumptions hold with $H_a^{\mathrm{RG}} = I$. The mixture is not log-concave when the means are sufficiently separated. Indeed, compare two phases, say 1 and 2, and write $\varphi$ for the standard Gaussian density in $E_C$. At $m_i$, the mixture density is at least $w_i\varphi(0)$, while at $z = (m_1 + m_2)/2$ it is at most $\varphi(0)e^{-B^2/4}$. For the displayed weights and $B \geq 4\sqrt{\log N}$, this is smaller than $\sqrt{w_1 w_2}\varphi(0)$ for every $N \geq 2$, contradicting the midpoint inequality required by log-concavity. Thus the example is genuinely non-log-concave.

Under $\nu_{C,1} = \mathcal{N}(m_1, I)$, the centered variables

$$Z_{1,(b,k)} = Y_{b,k} - \mathbb{E}_{\nu_{C,1}} Y_{b,k} = \langle e_{b,k}, G\rangle, \qquad G \sim N(0, I),$$

are independent standard Gaussians over $(b,k) \in \mathcal{T}_{\text{full}}$. Therefore the standard Gaussian maximum estimate gives

$$\mathbb{E} \max_{1\leq k\leq K} |G_{1,k}| \asymp \sqrt{\log K}, \qquad \mathbb{E} \max_{1\leq b\leq N,\, 1\leq k\leq K} |G_{b,k}| \asymp \sqrt{\log(NK)}.$$

These are exactly (A.10) and (A.11). Comparing these estimates with (A.8) and (A.9) proves the claimed strict separation. □

**Corollary A.7** (Pure-phase charts for a finite-volume double-well $\Phi^4$ model)**.** *Let $G = (V, E)$ be finite and*

$$S(\phi) = \frac{1}{2} \sum_{\{u,v\}\in E} c_{uv}(\phi_u - \phi_v)^2 + \lambda \sum_{u\in V} (\phi_u^2 - a^2)^2.$$

*Fix $m > a/\sqrt{3}$ and define phase charts*

$$\Omega_+ = \{\phi : \phi_u \geq m \ \forall u\}, \qquad \Omega_- = -\Omega_+.$$

*The restrictions of $e^{-S(\phi)}d\phi$ to $\Omega_+$ and $\Omega_-$ define two stable phase charts. A finite mixture of these chart measures is generally non-log-concave, but every chart has Hessian bounded below by*

$$L_G + 4\lambda(3m^2 - a^2)I.$$

*Therefore Theorem A.5 applies to every finite decomposition of variables $V = C \sqcup F$, and one nonlinear coarse-graining step preserves a two-chart pointwise-complexity envelope for these pure-phase restrictions. To cover the full double-well Gibbs measure one would need additional charts, for instance droplet or interface charts, and corresponding chartwise Hessian bounds.*

*Proof.* On either chart, $|\phi_u| \geq m$, and

$$\frac{d^2}{dt^2}\lambda(t^2 - a^2)^2 = 4\lambda(3t^2 - a^2) \geq 4\lambda(3m^2 - a^2) > 0.$$

Adding the graph Dirichlet Hessian gives the displayed lower bound. The result follows from Theorem A.5. □

**Interpretation of the non-convex phase-atlas example.** The finite nonconvex phase-atlas example should be read as a finite-cutoff bridge between convex concentration theory, high-dimensional geometry, and interacting-field intuition. Within a single stable phase chart, the measure is strongly log-concave: a uniform Hessian lower bound gives Brascamp–Lieb covariance domination and, through logarithmic-Sobolev/Herbst arguments, Gaussian-type concentration for linear and Lipschitz observables (Brascamp and Lieb, 1976; Bakry and Emery, 1985). A mixture of several such charts, however, is generally not log-concave and may be genuinely multimodal, so there need not be a single global convex potential or a single global Gaussian comparison. The phase-atlas decomposition restores tractability by separating these effects. Chartwise, the Schur-complement Hessian gives the local Gaussian comparison metric and hence the pointwise ellipsoidal complexity; globally, one pays an explicit atlas cost for selecting the relevant phase and active subspace.

This distinction is also natural from the viewpoint of convex geometry. Ordinary log-concavity gives powerful isoperimetric and tail estimates, and one-dimensional marginals of isotropic log-concave measures have universal sub-exponential tails (Borell, 1975; Lovasz and Vempala, 2007). It does not, however, by itself provide dimension-free Gaussian-type control in all directions. The Kannan–Lovasz–Simonovits conjecture is a central benchmark for dimension-free isoperimetric and Poincare-type control of isotropic log-concave measures (Kannan et al., 1995). The finite-cutoff viewpoint used here is different. Rather than trying to prove such global dimension-free control for an arbitrary log-concave measure, we work on local charts where the cutoff and the quadratic part of the effective action give a uniform curvature lower bound. On each chart this strong log-concavity yields Gaussian-type comparison in the anisotropic Hessian metric. Global non-log-concavity is then handled by the phase atlas, and directions below the finite resolution are paid for as approximation error rather than controlled through a global Cheeger constant. In this sense, the pointwise-complexity theorem gives an explicit spectral version of a local convex-geometric principle with global, finite-cutoff extension.

# B Graph Local-Time Application

## B.1 Local time versus cover and blancket time

We give a specialized application of the pointwise-complexity principle to the control of local times[2] in Gaussian free fields. The results separate two distinct ways of using Gaussian information, highlighting the advantage of a field-level route. In the field-level route, one fixes a target set, controls the pinned GFF on that set by a pointwise envelope, and then transfers the resulting estimate through the Ray–Knight theorem to a local-time statement. In the global route, one first establishes that the entire graph has been covered and only afterwards deduces that the target set has been visited. These two stopping-time arguments may occur at different scales.

We prove this separation on phase-star hierarchies. A target subatlas $I$ is encoded by an atlas prior $q$, and the same ambient pointwise envelope can be evaluated on the selected target. The exact law

$$\mathbb{P}\{\tau_I \le \tau(t)\} = (1 - e^{-t})^{|I|}, \qquad \mathbb{P}\{\tau_{\mathrm{cov}} \le \tau(t)\} = (1 - e^{-t})^{N}$$

shows that target cover has intrinsic inverse-root-local-time scale $\log |I|$, whereas any route through full cover pays $\log N$. The pointwise envelope pays the target code length $\log(|I|/q(I))$, and the chart-decorated phase-star adds a compressed within-chart term.

[2] Local time is a standard tool towards cutoff-uniform local integrability control in constructive QFT (Dynkin, 1984).

For readers coming from learning theory rather than probability, this is the graph version of the local/global separation. The selected target set $I$ plays the role of a pointwise subatlas; the full cover or blanket event plays the role of a global covering number over all phases. The similarity between pointwise/local dimension, covering number, cover time, and blanket time is only partially developed here, but the star examples make the basic scale separation explicit.

**Pinned GFF on graph, local time, and Ray–Knight.** Let $G = (V, E, c)$ be a finite connected weighted graph with symmetric conductances $c_{xy} = c_{yx} \geq 0$, where $c_{xy} > 0$ exactly when $\{x, y\} \in E$. Write

$$c_x := \sum_{y \in V} c_{xy}.$$

The weighted graph Laplacian acts on functions $f : V \to \mathbb{R}$ by

$$(Lf)(x) = \sum_{y} c_{xy}\big(f(x) - f(y)\big).$$

We consider the continuous-time random walk, started at a distinguished root $v_0$, that jumps from $x$ to $y$ at rate $c_{xy}/c_x$. Its normalized local time is

$$L_t(x) = \frac{1}{c_x}\int_0^t \mathbf{1}\{X_s = x\}\, ds, \qquad \tau(t) := \inf\{s \geq 0 : L_s(v_0) > t\}.$$

The pinned Gaussian free field with root $v_0$ is the centered Gaussian vector $\eta = \{\eta_x\}_{x \in V}$ such that $\eta_{v_0} = 0$ almost surely and whose covariance is the killed Green function

$$\mathbb{E}[\eta_x \eta_y] = \Gamma_{v_0}(x, y) := \mathbb{E}_x[L_{T_{v_0}}(y)],$$

where $T_{v_0}$ is the hitting time of $v_0$. Equivalently, this covariance is the inverse of the pinned Laplacian on $V \setminus \{v_0\}$. Its canonical metric is

$$d(x, y) = \big(\mathbb{E}|\eta_x - \eta_y|^2\big)^{1/2} = \sqrt{R_{\mathrm{eff}}(x, y)},$$

where $R_{\mathrm{eff}}$ is the effective-resistance metric of the conductance network.

We use the following finite-graph form of the second Ray–Knight theorem (Dynkin, 1984). Let $\eta'$ be an independent copy of the pinned GFF.

**Proposition B.1** (Second Ray–Knight isomorphism on a finite graph)**.** *With the above normalization, under the product law of the walk and the independent fields* $(\eta, \eta')$*,*

$$\Big(L_{\tau(t)}(x) + \tfrac{1}{2}\eta_x^2\Big)_{x \in V} \overset{d}{=} \Big(\tfrac{1}{2}(\eta'_x + \sqrt{2t})^2\Big)_{x \in V}. \tag{B.1}$$

*In particular,* $L_{\tau(t)}(v_0) = t$ *almost surely.*

**Cover and Blancket Time.** For a finite connected conductance graph $G = (V, E, c)$, let $\tau_{\mathrm{cov}}$ be the first time every vertex has been visited, and write

$$t_{\mathrm{cov}}(G) := \max_{v \in V} \mathbb{E}_v[\tau_{\mathrm{cov}}]. \tag{B.2}$$

A blanket time strengthens cover time by requiring that the walk has accumulated roughly stationary proportions of visits at all vertices. One convenient strong version leads to

$$t_{\mathrm{bl}}(G, \delta) := \max_{v \in V} \mathbb{E}_v[\tau_{\mathrm{bl}}(\delta)]. \tag{B.3}$$

**Main results on local-time envelopes and cover/blanket criteria.** Let $H \subseteq V$ be a target set. The proof below only requires a deterministic envelope for the pinned GFF on $H$. Such an envelope may be obtained either by applying Theorem 1.1 to the restricted field on $H \cup \{v_0\}$, or by applying it once to a larger ambient field and then evaluating the resulting bound on $H$. This distinction is useful in the phase-star examples, where one ambient atlas prior is fixed before the target subatlas is selected.

**Corollary B.2** (Target-set pointwise local-time envelope). *Let $A_H(\cdot;\delta)$ be any deterministic function satisfying*

$$\mathbb{P}\big\{|\eta_x| \le A_H(x;\delta) \text{ for all } x \in H\big\} \ge 1-\delta$$

*for the pinned GFF. Then, under the law of the walk, with probability at least $1-\delta$,*

$$\forall x \in H : \qquad \big[t - \sqrt{2t}\, A_H(x;\delta/2)\big]_+ \le L_{\tau(t)}(x) \le t + \sqrt{2t}\, A_H(x;\delta/2) + \frac{1}{2} A_H(x;\delta/2)^2. \tag{B.4}$$

Let

$$\tau_H := \inf\{s \ge 0 : L_s(x) > 0 \text{ for every } x \in H\}$$

be the target-cover time of $H$. The preceding envelope gives the following target-cover and target-blanket criterion.

**Corollary B.3** (Target cover and target blanket criterion). *Assume that for some $t > 0$, $H \subseteq V$, and $\theta \in (0, 1/\sqrt{2})$,*

$$\sup_{x\in H} A_H(x;\delta/2) \le \theta\sqrt{t}.$$

*Then, under the law of the walk, with probability at least $1-\delta$,*

$$(1-\sqrt{2}\,\theta)t \le L_{\tau(t)}(x) \le \Big(1 + \sqrt{2}\,\theta + \frac{1}{2}\theta^2\Big)t \qquad \forall x \in H.$$

*In particular, $\tau_H \le \tau(t)$ on this event, and*

$$\max_{u,v\in H} \frac{L_{\tau(t)}(u)}{L_{\tau(t)}(v)} \le \frac{1+\sqrt{2}\,\theta + \theta^2/2}{1-\sqrt{2}\,\theta}.$$

*Proof of Corollary B.2.* Set

$$A_x := A_H(x;\delta/2), \qquad U_x := L_{\tau(t)}(x) + \frac{1}{2}\eta_x^2, \qquad V_x := \frac{1}{2}(\eta'_x + \sqrt{2t})^2, \qquad x \in H.$$

By Theorem B.1, $(U_x)_{x\in H} \overset{d}{=} (V_x)_{x\in H}$. Define

$$D_x := \frac{1}{2}(\sqrt{2t} - A_x)_+^2, \qquad B_x := t + \sqrt{2t}\, A_x + \frac{1}{2}A_x^2.$$

By the envelope assumption applied to the independent copy $\eta'$, the event $\{|\eta'_x| \le A_x \text{ for all } x \in H\}$ has probability at least $1-\delta/2$. On this event, $D_x \le V_x \le B_x$ for all $x \in H$. Equality in distribution therefore gives

$$\mathbb{P}\{D_x \le U_x \le B_x \text{ for all } x \in H\} \ge 1-\delta/2.$$

Applying the same envelope to $\eta$ itself and taking a union bound, with probability at least $1-\delta$, we also have $|\eta_x| \le A_x$ for all $x \in H$. On the intersection of these events,

$$L_{\tau(t)}(x) = U_x - \frac{1}{2}\eta_x^2 \le B_x,$$

and

$$L_{\tau(t)}(x) \ge D_x - \frac{1}{2}A_x^2.$$

Using nonnegativity of local time and the identity

$$\max\left\{0, \frac{1}{2}(\sqrt{2t} - A_x)_+^2 - \frac{1}{2}A_x^2\right\} = \left[t - \sqrt{2t}\,A_x\right]_+$$

gives (B.4). The event in the conclusion depends only on the walk, so the same probability bound holds under the walk law. □

*Proof of Corollary B.3.* Under the displayed small-envelope assumption, Corollary B.2 gives, simultaneously for all $x \in H$,

$$L_{\tau(t)}(x) \ge (1 - \sqrt{2}\,\theta)t > 0$$

and

$$L_{\tau(t)}(x) \le \Big(1 + \sqrt{2}\,\theta + \frac{1}{2}\theta^2\Big)t.$$

Thus every vertex in $H$ has been visited by time $\tau(t)$, and the displayed ratio bound follows by dividing the upper estimate by the lower estimate. □

## B.2 Separating examples: target subatlas versus full cover

We now give a finite graph model that separates direct target-local-time control from a reduction through full cover time. If a deterministic target set is fixed in advance and one is allowed to redo the entire Gaussian analysis on that target alone, then classical generic chaining on the restricted field has the correct target scale. The point here is different. We fix one ambient atlas prior before the target is selected, and then evaluate the same pointwise envelope on any target subatlas. The target scale includes the code length of the selected subatlas under this prior. A proof that first establishes full cover loses this target dependence, because it must cover every phase.

**Theorem B.4** (Phase-star target subatlas versus full cover)**.** *Let $S_N$ be the star graph with root $v_0$, leaves $\{1, \ldots, N\}$, and unit conductances. Let $\mathcal{I}$ be a collection of nonempty subsets of $\{1, \ldots, N\}$, and let $q \in \Delta(\mathcal{I})$. For $I \in \mathcal{I}$, set*

$$H_I := I, \qquad \tau_I := \inf\{s \ge 0 : L_s(i) > 0 \text{ for every } i \in I\}.$$

*Let $\tau_{\text{cov}}$ be the cover time of the whole star. Define the ambient prior*

$$\mu_q := \frac{1}{2}\delta_{v_0} + \frac{1}{4}\text{Unif}\{1, \ldots, N\} + \frac{1}{4}\sum_{I \in \mathcal{I}} q(I)\,\text{Unif}(I). \tag{B.5}$$

*Let $A_q(i; \delta)$ be the two-sided pointwise Gaussian envelope from (1.5) applied to the full pinned GFF on $S_N$, with prior $\mu_q$, with $r_0 = s_0 = 1$, and with constants enlarged if necessary. Then the following hold.*

(i) *The pinned GFF satisfies $\eta_1, \ldots, \eta_N$ independent $N(0, 1)$. Moreover, with probability at least $1 - \delta$, simultaneously for all $I \in \mathcal{I}$,*

$$\sup_{i \in I} |\eta_i| \le C\left[\sqrt{\log\frac{4|I|}{q(I)}} + \sqrt{\log\left(\frac{2 + \log\log(N + 3)}{\delta}\right)}\right]. \tag{B.6}$$

*Consequently, the same deterministic right-hand side may be used as $\sup_{i \in I} A_q(i; \delta)$ in the Ray–Knight local-time envelope.*

(ii) *For every* $\theta \in (0, 1/\sqrt{2})$*, there is* $C_\theta < \infty$ *such that, for each* $I \in \mathcal{I}$*,*

$$t \geq C_\theta \left[ \log \frac{4|I|}{q(I)} + \log \left( \frac{2 + \log\log(N+3)}{\delta} \right) \right] \implies \mathbb{P}\{\tau_I \leq \tau(t)\} \geq 1 - \delta. \tag{B.7}$$

*The probability statement is for the selected target* $I$*. The prior* $\mu_q$ *and the Gaussian envelope are fixed independently of that selection.*

(iii) *The exact inverse-root-local-time laws are*

$$\mathbb{P}\{\tau_I \leq \tau(t)\} = (1 - e^{-t})^{|I|}, \qquad \mathbb{P}\{\tau_{\mathrm{cov}} \leq \tau(t)\} = (1 - e^{-t})^{N}. \tag{B.8}$$

*Thus the* $(1-\delta)$*-quantile for covering* $k$ *prescribed leaves is*

$$q_k(\delta) := -\log\left(1 - (1-\delta)^{1/k}\right) = \log k + O_\delta(1). \tag{B.9}$$

*In particular,* $q_N(\delta) - q_{|I|}(\delta) = \log(N/|I|) + O_\delta(1)$*. Hence, whenever*

$$\log \frac{|I|}{q(I)} = o(\log N),$$

*the direct target-local-time route has an inverse-root-local-time scale asymptotically smaller than the scale required by any argument that certifies the same target event only through full cover.*

*Proof.* For the star graph, the pinned Laplacian on the leaves is the identity matrix, so the pinned GFF has covariance $I_N$. For a leaf $i$, $\sigma(i) = d(i, v_0) = 1$, and the star metric satisfies $d(i,j) = \sqrt{2}$ for distinct leaves. If $i \in I$, then (B.5) gives

$$\mu_q(\{i\}) \geq \frac{q(I)}{4|I|}.$$

Since the metric diameter of the star leaves is bounded by an absolute constant, the localized Fernique–Talagrand functional satisfies

$$\Phi_{\mu_q}(i) \leq C\sqrt{\log \frac{4|I|}{q(I)}} \qquad (i \in I).$$

The uniform component gives $\mu_q(\{i\}) \geq 1/(4N)$ for every leaf. Hence $\Phi_{\mu_q,*} \leq C\sqrt{\log(N+1)}$. With $r_0 = s_0 = 1$, the peeling multiplicity in Theorem 1.1 is bounded by a constant times $2+\log\log(N+3)$. The two-sided pointwise Gaussian theorem gives (B.6). Applying Corollary B.3, with the usual confidence split, gives (B.7) after increasing $C_\theta$.

It remains to prove (B.8). During visits to the root, the walk waits an exponential time of rate one before departing and then chooses one of the $N$ leaves uniformly. Since $c_{v_0} = N$, accumulated root occupation time at $\tau(t)$ is $Nt$. Therefore, per unit normalized root local time, departures from the root form a Poisson process of rate $N$ with independent uniform leaf marks. By Poisson thinning, for each leaf $i$, the number of excursions from the root to $i$ before $\tau(t)$ is Poisson with mean $t$, and these counts are independent over leaves. A leaf has positive local time by $\tau(t)$ if and only if it has been chosen at least once. This gives the two identities in (B.8). Solving $(1-e^{-t})^k \geq 1-\delta$ gives (B.9), and the asymptotic formula follows from $(1-\delta)^{1/k} = 1 + k^{-1}\log(1-\delta) + o(k^{-1})$. □

The collapsed star has no within-phase geometry. The next theorem adds a local chart to each phase. It is a graph analogue of a phase/subspace atlas: the prior first pays for selecting a target subatlas, and the effective resistance inside each dense chart suppresses the within-chart fluctuation. This gives a concrete finite-graph analogue of the phase/subspace atlas cost in Appendix A.

**Theorem B.5** (Decorated phase-star subatlas theorem)**.** *Fix $N, K \geq 1$ and $\kappa > 0$ with $\kappa(K+1) \geq 1$. Let $G_{N,K,\kappa}$ be the conductance graph with root $v_0$, phase centers $c_1, \dots, c_N$, and local chart vertices $z_{a,1}, \dots, z_{a,K}$ for each phase $a$. The root is connected to each $c_a$ by an edge of conductance one. For each $a$,*

$$C_a := \{c_a, z_{a,1}, \dots, z_{a,K}\}$$

*is a complete graph with edge conductance $\kappa$, and there are no other edges. Let $\mathcal{I}$ be a collection of nonempty subsets of $\{1, \dots, N\}$, and let $q \in \Delta(\mathcal{I})$. For $I \in \mathcal{I}$, define the decorated target*

$$H_{I,K} := \bigcup_{a \in I} C_a.$$

*Define the ambient prior*

$$\mu_{q,K} := \frac{1}{2}\delta_{v_0} + \frac{1}{4}\mathrm{Unif}\left(\bigcup_{a=1}^{N} C_a\right) + \frac{1}{4}\sum_{I \in \mathcal{I}} q(I)\,\mathrm{Unif}\left(\bigcup_{a \in I} C_a\right). \tag{B.10}$$

*Then the following hold.*

(i) *The effective resistance within a phase is compressed:*

$$R_{\mathrm{eff}}(x, y) \leq \frac{2}{\kappa(K+1)}, \qquad x, y \in C_a, \tag{B.11}$$

*while*

$$R_{\mathrm{eff}}(v_0, c_a) = 1, \qquad R_{\mathrm{eff}}(v_0, x) \leq 1 + \frac{2}{\kappa(K+1)}, \quad x \in C_a. \tag{B.12}$$

(ii) *Let $A_{q,K}(x;\delta)$ be the two-sided pointwise Gaussian envelope from (1.5) applied to the full pinned GFF on $G_{N,K,\kappa}$, with prior $\mu_{q,K}$, with $r_0 = s_0 = 1$, and with constants enlarged if necessary. Then, with probability at least $1-\delta$, simultaneously for all $I \in \mathcal{I}$,*

$$\begin{aligned}\sup_{x \in H_{I,K}} |\eta_x| \leq C\Bigg[&\sqrt{\log \frac{4|I|}{q(I)}} + \sqrt{\frac{1}{\kappa(K+1)} \log \frac{4|I|(K+1)}{q(I)}} \\ &+ \sqrt{\log\left(\frac{2 + \log\log((N+3)(K+3))}{\delta}\right)}\Bigg].\end{aligned} \tag{B.13}$$

*Consequently, for every $\theta \in (0, 1/\sqrt{2})$, there is $C_\theta < \infty$ such that, for the selected $I$,*

$$\begin{aligned}t \geq C_\theta\Bigg[&\log \frac{4|I|}{q(I)} + \frac{1}{\kappa(K+1)} \log \frac{4|I|(K+1)}{q(I)} \\ &+ \log\left(\frac{2 + \log\log((N+3)(K+3))}{\delta}\right)\Bigg].\end{aligned} \tag{B.14}$$

*implies*

$$\mathbb{P}\{\tau_{H_{I,K}} \leq \tau(t)\} \geq 1 - \delta.$$

(iii) *Let* $\tau_{\rm cov}$ *be the cover time of the full decorated graph, and set* $C_N^{\rm cen} := \{c_1, \dots, c_N\}$*.*

$$\mathbb{P}\{C_N^{\rm cen} \subseteq X[0, \tau(t)]\} = (1 - e^{-t})^N. \tag{B.15}$$

*Since full cover implies that all phase centers have been visited,*

$$\mathbb{P}\{\tau_{\rm cov} \le \tau(t)\} \le (1 - e^{-t})^N. \tag{B.16}$$

*Thus any argument that proves target cover only by first proving full cover at confidence* $1-\delta$ *must use inverse-root-local-time at least* $q_N(\delta)$ *from* (B.9)*. If, for fixed* $\delta \in (0,1)$*,*

$$\log \frac{|I|}{q(I)} + \frac{1}{\kappa(K+1)} \log \frac{|I|(K+1)}{q(I)} + \log\log((N+3)(K+3)) = o(\log N),$$

*then the sufficient scale in* (B.14) *is* $o(q_N(\delta))$*, while* $\mathbb{P}\{\tau_{\rm cov} \le \tau(t)\} \to 0$ *along that scale.*

The decorated phase-star should be compared with the DNN and RG examples as follows. The centers $c_a$ are global phase labels, while each clique $C_a$ is a local chart. The envelope in (B.13) has a target-atlas code $\log(4|I|/q(I))$ and a compressed within-chart term $(\kappa(K+1))^{-1}\log(4|I|(K+1)/q(I))$. The latter is small when the clique conductance is large, just as irrelevant RG directions or compressed DNN feature directions are small below finite resolution. Any argument through full cover ignores this selected atlas and pays the coupon-collector scale $\log N$.

*Proof.* The resistance estimate inside $C_a$ is the standard effective resistance of a complete graph with $K+1$ vertices and edge conductance $\kappa$, namely $2/[\kappa(K+1)]$. Since $c_a$ is the only vertex of $C_a$ connected to the root, attaching the remaining clique to $c_a$ does not change the effective resistance between $v_0$ and $c_a$; hence $R_{\rm eff}(v_0, c_a) = 1$. The triangle inequality for effective resistance gives (B.12).

Put

$$\alpha_{K,\kappa} := \sqrt{\frac{2}{\kappa(K+1)}}.$$

For $x \in C_a$ and $a \in I$, the ball $B_d(x, \alpha_{K,\kappa})$ contains $C_a$ by (B.11). Under $\mu_{q,K}$, this ball has mass at least $q(I)/(4|I|)$. For smaller radii, the singleton mass is at least $q(I)/[4|I|(K+1)]$. Also $\sigma(x) \le 2$ by (B.12). Therefore, for $x \in C_a$ with $a \in I$,

$$\Phi_{\mu_{q,K}}(x) \le C\sqrt{\log \frac{4|I|}{q(I)}} + C\alpha_{K,\kappa}\sqrt{\log \frac{4|I|(K+1)}{q(I)}}.$$

The uniform component in $\mu_{q,K}$ gives singleton mass at least $1/[4N(K+1)]$, and hence $\Phi_{\mu_{q,K},*} \le C\sqrt{\log((N+1)(K+1))}$. With $r_0 = s_0 = 1$, the peeling multiplicity contributes only the displayed $\log\log((N+3)(K+3))$ term. This proves (B.13). The target-cover implication follows from Corollary B.3 as in the proof of Theorem B.4.

Finally, each departure from the root chooses one of the phase centers uniformly, and the same Poisson-thinning argument used in Theorem B.4 gives (B.15). Since full cover requires every phase center to have been visited, (B.16) follows. The final asymptotic statement follows from $q_N(\delta) = \log N + O_\delta(1)$. □

**Relation to classical cover-time theory.** Let $H(u,v)$ be the expected hitting time from $u$ to $v$, and define the commute-time metric

$$\kappa(u,v) := H(u,v) + H(v,u).$$

Let $\mathcal{C} := \sum_{x\in V} c_x$ be the total conductance. The commute-time identity gives

$$\kappa(u,v) = \mathcal{C}\, R_{\mathrm{eff}}(u,v).$$

For an unweighted graph, $\mathcal{C} = 2|E|$. If $\eta = \{\eta_v\}_{v\in V}$ denotes the pinned GFF, then $\mathbb{E}(\eta_u - \eta_v)^2 = R_{\mathrm{eff}}(u,v)$.

**Proposition B.6** (Ding–Lee–Peres (Ding et al., 2012))**.** *For any connected conductance graph $G = (V,E,c)$ and any fixed $0 < \delta < 1$,*

$$t_{\mathrm{cov}}(G) \;\asymp\; \gamma_2\big(V, \sqrt{\kappa}\big)^2 \;=\; \mathcal{C}\cdot\gamma_2\big(V, \sqrt{R_{\mathrm{eff}}}\big)^2 \;\asymp_\delta\; t_{\mathrm{bl}}(G,\delta). \tag{B.17}$$

*Equivalently,*

$$t_{\mathrm{cov}}(G) \;\asymp\; \mathcal{C}\cdot\Big(\mathbb{E}\max_{v\in V}\eta_v\Big)^2. \tag{B.18}$$

Proposition B.6 is a global theorem: it identifies the time scale needed to cover every vertex. The local-time estimates above are not meant to recover this general cover-time characterization. Rather, they address a different question by comparing local time and cover time through their defining events. They retain a single ambient field-level envelope and evaluate it on a selected subatlas before any full-cover event is invoked. Theorems B.4 and B.5 show that, on the phase-star example, this distinction can be strict. On the star, a selected target $I$ has inverse-root-local-time quantile $\log|I| + O_\delta(1)$, while full cover has quantile $\log N + O_\delta(1)$. The pointwise target envelope pays the atlas code length $\log(|I|/q(I))$; in the decorated star it also pays the compressed within-chart term $(\kappa(K+1))^{-1}\log(|I|(K+1)/q(I))$. Any argument that first passes through the full-cover event necessarily pays the coupon-collector cost associated with all $N$ phase centers. If a target set is fixed in advance, and the Gaussian analysis is allowed to be restricted to that target, then classical generic chaining applied to the restricted field can recover the corresponding target scale. The purpose of Theorems B.4 and B.5, however, is different: they construct a simultaneous envelope valid uniformly over all target subsets $I$. The pointwise formulation is what makes this possible. It retains the dependence on the selected subatlas under a single ambient prior, rather than collapsing the analysis at the outset to the full-cover event.

# C Technical variants and Sudakov-type consequences

In this section, we derive a Sudakov-type comparison from the Bayesian algorithmic lower bound for the nearest-neighbor decoder via a supremum relaxation. Our aim is conceptual: to show that the global Sudakov scale may be viewed as a coarse relaxation of a more localized algorithmic certificate. We use this route only to clarify the soundness and intuition of the approach. Recent concurrent work (Zadik, 2026) shows that a comparison between nearest-neighbor and Bayes-optimal risks yields a succinct proof of the majorizing-measure lower bound. Together with the standard upper-bound arguments, this gives a concise perspective on both sides of Talagrand's majorizing-measure theorem.

By replacing the exact ghost mass with a class-wide small-ball mass and optimizing this small-ball term over hard priors in Theorem 1.5, we obtain a nearest-neighbor route to a Sudakov-type comparison for priors localized at the tested radius.

**Corollary C.1** (Sudakov-type comparison through supremum relaxation)**.** *Under the setup of Theorem 1.5 with the ordinary nearest-neighbor decoder $\Omega \equiv 0$ and the centered reference choice $m = \bar{v}_\pi$, set*

$$q_\pi(\Delta) := \sup_{a\in T} \pi(B_d(a,\Delta)), \qquad L_\pi(\Delta) := \log \frac{1}{q_\pi(\Delta)}.$$

*Since $\rho_{\Delta,Q_{\tau,\bar{v}_\pi}}(\widehat{\Theta}_{\mathrm{nn}}) \le q_\pi(\Delta)$, if $L_\pi(\Delta)$ is larger than a universal constant, one may choose $\tau^2 \asymp \mathsf{V}_\pi/L_\pi(\Delta)$ so that (1.15) holds with an absolute $c \in (0,1)$. Relaxing the nearest-neighbor lower bound (1.17) to supremum over the Gaussian process yields the variance-normalized single-radius obstruction*

$$\mathbb{E} \sup_{s,u\in T} (X_s - X_u) \gtrsim \Delta^2 \sqrt{\frac{L_\pi(\Delta)}{\mathsf{V}_\pi}}. \tag{C.1}$$

*In particular, if the hard prior is localized at scale $\Delta$, meaning $\mathsf{V}_\pi \le C_0\Delta^2$, then*

$$\mathbb{E} \sup_{s,u\in T} (X_s - X_u) \gtrsim_{C_0} \Delta\sqrt{L_\pi(\Delta)}. \tag{C.2}$$

*Equivalently, with*

$$N^{\mathrm{loc}}_{\mathrm{spread}}(T,d,\Delta;C_0) := \sup_{\pi:\,\mathsf{V}_\pi\le C_0\Delta^2} \frac{1}{\sup_{a\in T}\pi(B_d(a,\Delta))}, \tag{C.3}$$

*one obtains, whenever the logarithm is above the universal threshold required by the information condition,*

$$\mathbb{E} \sup_{s,u\in T} (X_s - X_u) \gtrsim_{C_0} \Delta\sqrt{\log N^{\mathrm{loc}}_{\mathrm{spread}}(T,d,\Delta;C_0)}. \tag{C.4}$$

*Proof of Corollary C.1.* For the variance-normalized small-ball statement, use the centered reference choice $m = \bar{v}_\pi$, for which $\mathcal{I}_{\pi,m} = \mathsf{V}_\pi/(4\tau^2)$, and the bound $\rho_{\Delta,Q_{\tau,\bar{v}_\pi}}(\widehat{\Theta}_{\mathrm{nn}}) \le q_\pi(\Delta)$. If $L_\pi(\Delta)$ is above a universal constant, choosing $\tau^2 \asymp \mathsf{V}_\pi/L_\pi(\Delta)$ with a sufficiently large absolute constant makes (1.15) hold for some absolute $c \in (0,1)$. Since $X_{\widehat{\Theta}_{\mathrm{nn}}} - X_\Theta \le \sup_{s,u\in T}(X_s - X_u)$ pointwise in $Z$, (1.17) gives

$$\mathbb{E} \sup_{s,u\in T} (X_s - X_u) \gtrsim \Delta^2 \sqrt{\frac{L_\pi(\Delta)}{\mathsf{V}_\pi}}.$$

If $\mathsf{V}_\pi \le C_0\Delta^2$, this becomes $\Delta\sqrt{L_\pi(\Delta)}$, with a constant depending only on $C_0$. Optimizing over such localized priors gives (C.4). □

The preceding corollary directly recovers a Sudakov-scale lower bound under a natural localized hard-prior condition. This is the precise mild condition needed by the nearest-neighbor route: the prior used to certify small balls must also have second moment comparable to the tested radius.

**Corollary C.2** (Localized packing form)**.** *Let $S \subset T$ be $2\Delta$-separated and assume that, for some $t_0 \in T$ and $C_0 < \infty$,*

$$S \subset B_d(t_0, C_0\Delta).$$

*If $|S| = M \ge M_0$, where $M_0$ is a universal constant, then*

$$\mathbb{E} \sup_{s,u\in T} (X_s - X_u) \gtrsim_{C_0} \Delta\sqrt{\log M}.$$

*Proof of Corollary C.2.* Apply Theorem 1.5 with $\pi$ equal to the uniform distribution on $S$. Since $S$ is $2\Delta$-separated, every $\Delta$-ball contains at most one point of $S$, so

$$q_\pi(\Delta) = \sup_{a \in T} \pi(B_d(a, \Delta)) \le M^{-1}, \qquad L_\pi(\Delta) \ge \log M.$$

The localization assumption gives

$$\mathsf{V}_\pi = \iint d(s,t)^2 \, \pi(ds)\pi(dt) \le 4C_0^2\Delta^2.$$

Equation (C.2) yields the claim. □

For comparison with classical covering notation, we record the finite-set fractional-covering duality used in Chen et al. (2024). Recall that the fractional covering number is equivalent to the canonical covering number, up to absolute-constant changes in the covering radius (Block et al., 2021; Chen et al., 2024).

**Lemma C.3** (Fractional-covering duality)**.** *For a finite metric space* $(T, d)$ *and any* $\Delta > 0$*,*

$$N_{\text{frac}}(T, d, \Delta) := \inf_{\mu \in \Delta(T)} \sup_{x \in T} \frac{1}{\mu(B_d(x, \Delta))} = \sup_{\pi \in \Delta(T)} \frac{1}{\sup_{a \in T} \pi(B_d(a, \Delta))}. \tag{C.5}$$

*Proof of Lemma C.3.* Let

$$r_* := \sup_{\mu \in \Delta(T)} \inf_{x \in T} \mu(B_d(x, \Delta)).$$

Then

$$N_{\text{frac}}(T, d, \Delta) = \frac{1}{r_*}.$$

Since $T$ is finite, von Neumann's minimax theorem gives

$$\begin{aligned} r_* &= \sup_{\mu \in \Delta(T)} \inf_{x \in T} \sum_{a \in T} \mu(a) \mathbf{1}\{a \in B_d(x, \Delta)\} \\ &= \inf_{\pi \in \Delta(T)} \sup_{\mu \in \Delta(T)} \sum_{x, a \in T} \pi(x)\mu(a) \mathbf{1}\{a \in B_d(x, \Delta)\} \\ &= \inf_{\pi \in \Delta(T)} \sup_{a \in T} \pi(B_d(a, \Delta)). \end{aligned}$$

Taking reciprocals yields (C.5). □

The equality (C.5) is an entropy duality. It optimizes the small-ball term, but it does not remove the factor $\mathsf{V}_\pi$ in (C.1), which comes from the information radius of the Gaussian location channel. There is therefore no contradiction between the fractional-covering lower bounds of Chen et al. (2024) and the localization requirement above: the fractional-covering duality identifies the optimal hard prior for the small-ball obstruction, while the nearest-neighbor Gaussian comparison also requires that this hard prior have controlled second moment at scale $\Delta$. Thus the comparison directly gives the localized obstruction (C.4) and the localized packing form in Corollary C.2.

The unrestricted classical Sudakov minoration is a separate Gaussian comparison theorem, not a consequence of linear-programming duality alone. Standard proofs use nontrivial Gaussian comparison and convex-geometric duality arguments; see, for example, the expositions of Pollard (2001) and Ledoux and Talagrand (1991, Chapter 3). The full lower half of the Fernique–Talagrand majorizing-measure theorem is stronger and multiscale (Talagrand, 2005; van Handel, 2018; Zadik, 2026). In the present paper we therefore keep the roles separate: the Bayesian algorithmic envelope supplies a one-scale, decision-theoretic, and prior-agnostic primitive, while the unrestricted Sudakov and generic-chaining lower bounds require the classical Gaussian comparison or multiscale localization machinery.

## References


Michael Aizenman and Hugo Duminil-Copin. Marginal triviality of the scaling limits of critical 4D Ising and $\phi_4^4$ models. *Annals of Mathematics*, 194(1):163–235, 2021.

Dominique Bakry and Michel Emery. Diffusions hypercontractives. In *Séminaire de probabilités XIX*, Lecture Notes in Mathematics, vol. 1123, pages 177–206. Springer, 1985.

Roland Bauerschmidt, David C. Brydges, and Gordon Slade. Introduction to a Renormalisation Group Method. *Lecture Notes in Mathematics*, vol. 2242. Springer, 2019.

Mikhail Belkin, Daniel Hsu, and Partha P. Mitra. Overfitting or perfect fitting? Risk bounds for classification and regression rules that interpolate. In *Advances in Neural Information Processing Systems 31*, 2018.

Adam Block, Yuval Dagan, and Alexander Rakhlin. Majorizing measures, sequential complexities, and online learning. In *Proceedings of the 34th Annual Conference on Learning Theory*, vol. 134, pages 1–4, 2021.

Christer Borell. Convex set functions in $d$-space. *Periodica Mathematica Hungarica*, 6:111–136, 1975.

Herm Jan Brascamp and Elliott H. Lieb. On extensions of the Brunn–Minkowski and Prekopa–Leindler theorems, including inequalities for log-concave functions, and with an application to the diffusion equation. *Journal of Functional Analysis*, 22(4):366–389, 1976.

David C. Brydges and Gordon Slade. A renormalisation group method. V. A single renormalisation group step. *Journal of Statistical Physics*, 159:589–667, 2015.

Sourav Chatterjee. Yang–Mills for probabilists. arXiv:1803.01950, 2018.

Fan Chen, Dylan J. Foster, Yanjun Han, Jian Qian, Alexander Rakhlin, and Yunbei Xu. Assouad, Fano, and Le Cam with interaction: A unifying lower bound framework and characterization for bandit learnability. arXiv:2410.05117, 2024.

Thomas M. Cover and Peter E. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.

George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2:303–314, 1989.

Jian Ding, James R. Lee, and Yuval Peres. Cover times, blanket times, and majorizing measures. *Annals of Mathematics*, 175(3):1409–1471, 2012.

E. B. Dynkin. Local times and quantum fields. In *Seminar on Stochastic Processes, 1983*. Birkhäuser, 1984.

Xavier Fernique. Régularité des trajectoires des fonctions aléatoires gaussiennes. In *École d'Été de Probabilités de Saint-Flour IV–1974*, Lecture Notes in Mathematics, vol. 480. Springer, 1975.

Dylan J. Foster, Sham M. Kakade, Jian Qian, and Alexander Rakhlin. The statistical complexity of interactive decision making. arXiv:2112.13487, 2021.

Dylan J. Foster, Noah Golowich, and Yanjun Han. Tight guarantees for interactive decision making with the decision-estimation coefficient. In *Proceedings of the 36th Annual Conference on Learning Theory*, vol. 195, pages 1–75, 2023.

Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.

Arthur Jaffe and Edward Witten. Quantum Yang–Mills theory. Clay Mathematics Institute Millennium Prize Problem description, 2000.

Ravi Kannan, Laszlo Lovasz, and Miklos Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry*, 13:541–559, 1995.

Shaojie Li and Yunbei Xu. Pointwise generalization in deep neural networks. arXiv:2605.18598, 2026.

Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. Springer, 1991.

Leonid A. Levin. Average case complete problems. *SIAM Journal on Computing*, 15(1):285–286, 1986.

Laszlo Lovasz and Santosh Vempala. The geometry of logconcave functions and sampling algorithms. *Random Structures & Algorithms*, 30(3):307–358, 2007.

Neil Lutz. A note on pointwise dimensions. arXiv:1612.05849, 2016.

David Pollard. The duality proof of the Sudakov minoration. Lecture notes, Yale University, available at http://www.stat.yale.edu/~pollard/Notes/Sudakov.pdf, 2001.

Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM*, 51(3):385–463, 2004.

Michel Talagrand. *The Generic Chaining: Upper and Lower Bounds of Stochastic Processes*. Springer, 2005.

Michel Talagrand. Regularity of Gaussian processes. *Acta Mathematica*, 159:99–149, 1987.

Ramon van Handel. Chaining, interpolation and convexity II: The contraction principle. *The Annals of Probability*, 46(3):1764–1805, 2018.

Yunbei Xu and Assaf Zeevi. Towards optimal problem dependent generalization error bounds in statistical learning theory. *Mathematics of Operations Research*, 2025.

Yunbei Xu and Assaf Zeevi. Bayesian design principles for frequentist sequential learning. *Journal of the ACM*, 2025.

Andrew C. Yao. Probabilistic computations: Toward a unified measure of complexity. In *Proceedings of the 18th Annual Symposium on Foundations of Computer Science*, pages 222–227, 1977.

Ilias Zadik. A Bayesian proof and interpretation of Talagrand's majorizing measure theorem. *arXiv preprint arXiv:2605.30321*, 2026.

Tong Zhang. Information theoretical upper and lower bounds for statistical estimation. *IEEE Transactions on Information Theory*, 52(4):1307–1321, 2006.