Algorithmic Complexity in Gaussian-Process Bandits

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

Abstract

Gaussian-process upper confidence bound (GP-UCB) is a central practical algorithm in Gaussian-process bandits, Bayesian optimization, operations research, and sequential AI. Whether standard GP-UCB is minimax optimal has been debated for more than a decade. This paper does not attempt to resolve that debate. Instead, it gives a structural explanation of the classical GP-UCB proof in the language of algorithmic information, and establishes separations that challenge full-class minimax criteria.

The main observation is that GP-UCB is naturally interpreted through an *algorithmic prior*: the surrogate GP prior need not be true in the frequentist RKHS setting, but it defines the posterior mean, posterior variance, and model-index information used by the algorithm. Frequentist validity is supplied by a calibration theorem. Using a Gaussian-process algorithmic prior, we apply the model-index algorithmic information ratio (MAIR) theorem of Xu and Zeevi (2025) to obtain a regret bound for *arbitrary* algorithms. Specializing to GP-UCB, we compute the MAIR gradient under the surrogate GP belief and verify the resulting error condition using the two classical ingredients of GP-UCB: RKHS self-normalized calibration and optimism. This recovers the realized-information bound $R_T \lesssim \left[B + \frac{R}{\sqrt{\lambda}}\sqrt{\mathcal{I}_T + \log(1/\delta)}\right]\sqrt{T\mathcal{I}_T}$, $\mathcal{I}_T = \frac{1}{2}\log\det(I + \lambda^{-1}K_T)$, with maximal information gain appearing only after the worst-case relaxation ($\mathcal{I}_T \le \gamma_T$).

Thus maximal information gain is not the intrinsic algorithmic quantity. It upper-bounds two distinct objects: cumulative algorithmic information gain and the MAIR/AIR coefficient inherited from self-normalized calibration. The COLT open problem on tight online RKHS confidence intervals is precisely about improving the second use. We isolate a general coefficient replacement based on effective ratios of one-step mean-shift operators, and show how it can be controlled by low-rank approximation, inducing points, random features, or regularized nearest-neighbor projections.

Finally, a finite hub–cloud GP-bandit construction shows that algorithmic complexity can be more informative than class-wide minimax or DEC certificates in overparameterized models. A full-action, complexity-weighted GP-UCB rule uses an agnostic complexity prior and achieves hub-scale performance under hub truths, while the full ambient class has cloud-scale minimax and DEC-style lower bounds. The separation occurs within a fixed DMSO/GP-bandit model and on the AIR/coefficient side, not through Bellman-rank representation vacuity or estimation-complexity bookkeeping.

Contents

1 Introduction

Gaussian-process bandits are a central mathematical model for bandit optimization: an algorithm sequentially evaluates an expensive black-box function and must trade off exploitation and exploration. GP-UCB and its variants are especially influential because they are simple, robust, and easy to implement in operations research and AI applications. The original GP-UCB paper of Srinivas et al. (2010) connected no-regret Bayesian optimization with experimental design through information gain. The frequentist RKHS analysis of Chowdhury and Gopalan (2017) then gave a powerful kernelized bandit theory using self-normalized Hilbert-space martingales.

The central scalar parameter in this theory is the maximal information gain

$$\gamma_T = \sup_{A \subseteq \mathcal{X},\ |A|=T} I(\widehat{f}_A; Y_A),$$

where $\widehat{f}$ is the surrogate GP. The worst-case nature of γ_T is useful for clean statements, but it hides two different roles. First, γ_T upper-bounds the information actually collected by the realized design. Second, through the known online RKHS confidence theorem, it appears inside the confidence-width coefficient. The first role is usually unavoidable at the level of class-wide information accumulation; the second is the suspicious extra factor.

This distinction is exactly the content of the COLT open problem of Vakili et al. (2021a). They formulate online RKHS confidence intervals and recall the Chowdhury–Gopalan bound

$$|f(x) - \mu_n(x)| \le \rho_n(\delta)\sigma_n(x), \qquad \rho_n(\delta) = B + R\sqrt{2\{\gamma_{n-1} + 1 + \log(1/\delta)\}}.$$

They ask whether the width can be reduced by an $\widetilde{O}(\sqrt{\gamma_n})$ factor. Plugging the known width into GP-UCB gives $\widetilde{O}(\gamma_T\sqrt{T})$ regret, which may fail to be sublinear for Matérn kernels; the desired shape is closer to $\widetilde{O}(\sqrt{dT\gamma_T})$ (Vakili et al., 2021a).

This paper does not claim to settle the minimax-optimality debate for GP-UCB. The debate is active. The minimax lower bounds of Scarlett et al. (2017) show that information-theoretic barriers are real for standard kernels. Vakili et al. (2021b) improved information-gain bounds and closed large polynomial gaps for Matérn kernels at the level of γ_T. Whitehouse et al. (2023) prove nearly optimal sublinear GP-UCB regret under eigendecay and smoothness-aware regularization. Bayesian analyses of GP-UCB and GP-TS have also improved, including realized-input-sequence and local-information arguments (Iwazaki, 2025; Takeno and Iwazaki, 2026). On the other hand, Lattimore (2023) gives lower-bound evidence that fully general adaptive confidence intervals cannot be arbitrarily tightened, and Wang and Zhang (2026) gives a Matérn lower-bound obstruction for GP-UCB under polynomial effective optimism.

Our contribution is conceptual and structural, but mathematically exact. We show that the frequentist GP-UCB proof is a model-index AIR proof. The surrogate GP prior is algorithmic, not true. Its posterior update defines a model-index information increment

$$g_t = I_{\rho_{t-1}}(\widehat{f}; Y_t \mid H_{t-1}, x_t) = \frac{1}{2}\log\left(1 + \lambda^{-1}\sigma^2_{t-1}(x_t)\right),$$

where ρ_{t-1} is the surrogate posterior. The cumulative quantity

$$\mathcal{I}_T = \sum_{t=1}^{T} g_t = \frac{1}{2} \log\det(I + \lambda^{-1} K_{x_{1:T}})$$

is the *algorithmic information gain* of the actual random GP-UCB trajectory. Classical proofs already pass through $\mathcal{I}_T$ and then relax to γ_T.

The more interesting point is the coefficient. Applying the MAIR post-validation theorem of Xu and Zeevi (2025) to the actual GP-UCB policy gives a closed-form gradient bracket. That bracket is controlled by exactly the two classical GP-UCB ingredients: self-normalized calibration and optimism. The result is a frequentist MAIR proof of the known realized-information regret bound. In other words,

$$\text{self-normalized calibration} + \text{UCB optimism} \quad \Longrightarrow \quad \text{MAIR gradient error condition.}$$

This explains why Bayesian Thompson-sampling information-ratio proofs and frequentist UCB proofs have long looked so similar: both are regret–information offset arguments. The Bayesian proof obtains calibration from the true posterior. The frequentist GP-UCB proof obtains calibration from a self-normalized martingale theorem.

The broader philosophy is algorithmic complexity. For a practical algorithm, the useful complexity is not necessarily the full-class minimax or DEC complexity. It is the information and coefficient of the algorithm's own decision law, with the surrogate prior understood as an algorithmic prior. We illustrate this by a finite hub–cloud GP bandit. A full-action, complexity-weighted GP-UCB algorithm has no oracle knowledge of the hub; it simply uses a global complexity prior or norm penalty. Under hub truths it never wastes queries on a huge irrelevant cloud and has hub-scale regret, while the full ambient class contains cloud truths forcing cloud-scale minimax and DEC-style lower bounds. This separation does not say that minimax or DEC is wrong. It says that overparameterized model classes often contain many hypotheses that a practical regularized algorithm intentionally refuses to be robust to. In such settings, algorithmic MAIR/AIR is the right certificate for the algorithm, while minimax/DEC is the right certificate for a different, class-wide robustness question.

Contributions. The paper makes four points.

1. We formulate the realized-information GP-UCB proof as a model-index AIR proof. The proof applies the MAIR gradient theorem to *arbitrary* algorithms, computes the gradient, and verifies the error condition through self-normalized calibration and UCB optimism.

2. We separate the two information-gain roles: cumulative algorithmic information gain $\mathcal{I}_T$ and the AIR/MAIR coefficient. Both are algorithmic quantities, rather than relaxations based on the common maximal information gain γ_T.

3. We give a general effective-ratio coefficient theorem. When one-step regret and information are controlled by nuclear and Frobenius norms of a mean-shift operator, the AIR coefficient is an approximation-aware effective rank.

4. We give a full-action hub–cloud GP-bandit separation showing that algorithmic complexity can be much smaller than full-class minimax/DEC in overparameterized models.

2 Preliminaries: GP-UCB and realized kernel information

Let $(\mathcal{X}, k)$ be a finite action kernel bandit; compact domains may be handled by standard discretization, but finiteness keeps the MAIR formulas uncluttered. Let $\mathcal{H}_k$ be the RKHS and assume

$$\|f^\star\|_{\mathcal{H}_k} \le B, \qquad k(x,x) \le \kappa^2 \quad \forall x \in \mathcal{X}.$$

The learner observes

$$Y_t = f^\star(x_t) + \varepsilon_t,$$

where x_t is H_{t-1}-measurable and ε_t is conditionally R-sub-Gaussian. Write $\phi(x) = k(x,\cdot)$ and

$$V_t = \lambda I + \sum_{s=1}^{t} \phi(x_s) \otimes \phi(x_s), \qquad S_t = \sum_{s=1}^{t} \varepsilon_s \phi(x_s).$$

The surrogate GP posterior mean and variance are

$$\mu_t(x) = \left\langle \phi(x), V_t^{-1} \sum_{s=1}^{t} Y_s \phi(x_s) \right\rangle, \qquad \sigma_t^2(x) = \lambda \langle \phi(x), V_t^{-1} \phi(x) \rangle.$$

The posterior is algorithmic: $f^\star$ is not assumed to be drawn from it.

The realized kernel information gain is

$$\mathcal{I}_T^{\mathrm{ker}} := \frac{1}{2} \log\det(I + \lambda^{-1} K_T),$$

where $K_T = (k(x_i, x_j))_{i,j \le T}$. The determinant identity gives

$$\mathcal{I}_T^{\mathrm{ker}} = \frac{1}{2} \sum_{t=1}^{T} \log\left(1 + \lambda^{-1} \sigma_{t-1}^2(x_t)\right). \tag{1}$$

The usual maximal information gain is the worst-case relaxation

$$\gamma_T = \sup_{A \subseteq \mathcal{X},\ |A| = T} \frac{1}{2} \log\det(I + \lambda^{-1} K_A), \qquad \mathcal{I}_T^{\mathrm{ker}} \le \gamma_T.$$

The realized form (1) is already present in the variance-information lemmas of Srinivas et al. (2010) and Chowdhury and Gopalan (2017); replacing it by γ_T is a deterministic simplification, not an intrinsic step.

3 MAIR regret bounds for *arbitrary* algorithms with GP algorithmic priors

We recall the model-index version of algorithmic information ratio in a form tailored to kernelized bandits. Let $\mathcal{M}$ be a dominated model class. Under model M, decision π gives reward $r_M(\pi)$ and observation distribution $P_{M,\pi}$. For a decision distribution $p \in \Delta(\Pi)$, define

$$\Delta_M(p) := r_M(\pi_M) - \mathbb{E}_{\pi \sim p} r_M(\pi),$$

where π_M is an optimal decision under M. Let $\rho \in \Delta(\mathcal{M})$ be a reference belief and $\mu \in \Delta(\mathcal{M})$ an algorithmic belief. The predictive observation law under μ is

$$\mu_{o|\pi} := \int P_{N,\pi}\, \mu(dN).$$

Definition 3.1 (Model-index AIR). *For $\eta > 0$, the model-index AIR objective is*

$$\mathsf{MAIR}_{\rho,\eta}(p,\mu) := \mathbb{E}_{M\sim\mu} \Delta_M(p) - \frac{1}{\eta} \mathbb{E}_{\pi \sim p} I_\mu(M; O \mid \pi) - \frac{1}{\eta} \mathrm{KL}(\mu \| \rho),$$

where

$$I_\mu(M; O \mid \pi) = \mathbb{E}_{M\sim\mu} \mathrm{KL}(P_{M,\pi} \| \mu_{o|\pi}).$$

This can be viewed as the model-index analogue of the decision-marginal AIR of Russo and Van Roy (2014) and Xu and Zeevi (2025). For details, see Section 7 of Xu and Zeevi (2025). It is closer to GP-bandit analysis because its information increment is the function/model posterior-update KL. Ordinary decision AIR uses information about the optimal decision π_M; MAIR uses information about M itself. Since π_M is a function of M, decision information is at most model information.

Theorem 3.2 (Restatement of Xu–Zeevi Theorem 7.1: MAIR regret bound)**.** *Consider the stochastic setting with a locally compact model class $\mathcal{M}$ and a fixed ground-truth model $M^\star \in \mathcal{M}$. Let an arbitrary algorithm produce decision distributions $p_1, \dots, p_T$, and let $\mu_1, \dots, \mu_T$ be any sequence of algorithmic beliefs. Let $\rho_t \in \operatorname{int}(\Delta(\mathcal{M}))$ denote the round-t reference belief, generated by the usual posterior update from the previous algorithmic belief, $\rho_{t+1}(\cdot) = \mu_t(\cdot \mid \pi_t, o_t)$. Then, for every $\eta > 0$,*

$$R_T \leq \frac{1}{\eta}\mathbb{E}\left[\log \frac{\rho_{T+1}(M^\star)}{\rho_1(M^\star)}\right] + \mathbb{E}\left[\sum_{t=1}^{T}\left\{\mathsf{MAIR}_{\rho_t,\eta}(p_t,\mu_t) + \left\langle \frac{\partial}{\partial\mu}\mathsf{MAIR}_{\rho_t,\eta}(p_t,\mu)\Big|_{\mu=\mu_t}, \mathbf{1}(M^\star) - \mu_t \right\rangle\right\}\right],$$

where $\mathbf{1}(M^\star)$ denotes the vector in $\Delta(\mathcal{M})$ whose $M^\star$ coordinate is one and whose other coordinates are zero.

Major novelty: GP algorithmic priors in place of uniform finite priors. In the original finite theorem of Xu and Zeevi (2025), the term $\log|\mathcal{M}|/\eta$ is the price of a posterior-ratio telescope under a static finite model reference and a uniform algorithmic prior ρ_1. For GP-UCB this static-prior interpretation is not the right object: the GP prior is a surrogate algorithmic prior, while $f^\star$ is fixed in the RKHS. We therefore use the theorem in posterior-reference form: conditionally on H_{t-1}, set the reference belief to the current surrogate posterior ρ_{t-1} and keep the one-step information

$$g_t^{\text{alg}} := I_{\rho_{t-1}}(\widehat{f}; Y_t \mid H_{t-1}, x_t)$$

instead of replacing it by a finite prior-mass potential.

For the Gaussian observation model with noise variance λ,

$$g_t^{\text{alg}} = \frac{1}{2}\log\left(1 + \lambda^{-1}\sigma_{t-1}^2(x_t)\right), \qquad \sum_{t=1}^{T} g_t^{\text{alg}} = \frac{1}{2}\log\det(I + \lambda^{-1}K_T) = \mathcal{I}_T^{\text{ker}}.$$

This is only an algorithmic information-accounting device; frequentist validity for the fixed truth is still supplied by RKHS calibration and UCB optimism. The Gaussian algorithmic prior has the additional advantage that, for Gaussian-process bandits, posterior updates remain within a single Gaussian exponential family.

Lemma 3.3 (Frequentist evaluation of the GP posterior telescope)**.** *Under the setup of Section 2, let $q_1 = \mathsf{GP}(0, k)$ be the Gaussian-process reference prior. For each $\lambda > 0$, define the algorithmic posterior path q_t^λ by the Gaussian reference likelihood*

$$Y_s = f(x_s) + \xi_s, \qquad \xi_s \sim N(0, \lambda).$$

Then the following pathwise identity holds:

$$\log \frac{dq_{T+1}}{dq_1}(f^\star) = \mathcal{I}_T^{\text{ker}} + \frac{1}{2}Y_T^\top(K_T + \lambda I_T)^{-1}Y_T - \frac{1}{2\lambda}\|Y_T - f_T^\star\|_2^2.$$

Consequently,

$$\log \frac{dq_{T+1}}{dq_1}(f^\star) \leq \mathcal{I}_T^{\text{ker}} + \frac{1}{2}\|f^\star\|_{\mathcal{H}_k}^2.$$

In particular, under any fixed frequentist law $\mathbb{P}_{f^\star}$ for the errors and the algorithmic randomization,

$$\mathbb{E}_{f^\star}\left[\log \frac{dq_{T+1}}{dq_1}(f^\star)\right] \le \mathbb{E}_{f^\star}\left[\mathcal{I}_T^{\mathrm{ker}}\right] + \frac{1}{2}\|f^\star\|_{\mathcal{H}_k}^2.$$

If the design $x_{1:T}$ is nonadaptive, then $\mathcal{I}_T^{\mathrm{ker}}$ is deterministic and the first term on the right is simply $\mathcal{I}_T^{\mathrm{ker}}$.

Proof. The action-selection rule is the same under the fixed-truth model and under the GP reference model. Hence the action probabilities cancel in the likelihood ratio, and it is enough to compare the conditional observation likelihoods along the realized action path. Using the canonical likelihood-ratio version of the Radon–Nikodym derivative,

$$\frac{dq_{T+1}}{dq_1}(f) = \frac{\prod_{t=1}^T \varphi_\lambda(Y_t - f(x_t))}{\int \prod_{t=1}^T \varphi_\lambda(Y_t - g(x_t))\, q_1(dg)},$$

where φ_λ is the $N(0,\lambda)$ density. By Bayes' rule, the displayed Radon–Nikodym derivative is the posterior/prior density ratio associated with the Gaussian pseudo-likelihood.

For the realized design $x_{1:T}$, the GP reference marginal law of Y_T is

$$Y_T \sim N(0, K_T + \lambda I_T).$$

Therefore

$$\log \frac{dq_{T+1}}{dq_1}(f^\star) = \log \frac{(2\pi\lambda)^{-T/2}\exp\left(-\frac{1}{2\lambda}\|Y_T - f_T^\star\|_2^2\right)}{(2\pi)^{-T/2}\det(K_T+\lambda I_T)^{-1/2}\exp\left(-\frac{1}{2}Y_T^\top (K_T+\lambda I_T)^{-1}Y_T\right)}.$$

Since

$$\frac{1}{2}\log\frac{\det(K_T+\lambda I_T)}{\lambda^T} = \frac{1}{2}\log\det(I_T + \lambda^{-1}K_T) = \mathcal{I}_T^{\mathrm{ker}},$$

this gives the identity

$$\log \frac{dq_{T+1}}{dq_1}(f^\star) = \mathcal{I}_T^{\mathrm{ker}} + \frac{1}{2}Y_T^\top (K_T+\lambda I_T)^{-1}Y_T - \frac{1}{2\lambda}\|Y_T - f_T^\star\|_2^2.$$

It remains to upper-bound the quadratic term. We use the standard RKHS variational identity

$$Y_T^\top (K_T+\lambda I_T)^{-1}Y_T = \inf_{h\in\mathcal{H}_k}\left\{\|h\|_{\mathcal{H}_k}^2 + \frac{1}{\lambda}\|Y_T - h_T\|_2^2\right\},$$

where $h_T = (h(x_1),\dots,h(x_T))^\top$. Choosing $h = f^\star$ gives

$$Y_T^\top (K_T+\lambda I_T)^{-1}Y_T \le \|f^\star\|_{\mathcal{H}_k}^2 + \frac{1}{\lambda}\|Y_T - f_T^\star\|_2^2.$$

Substituting this bound into the previous identity cancels the noise quadratic term and yields

$$\log \frac{dq_{T+1}}{dq_1}(f^\star) \le \mathcal{I}_T^{\mathrm{ker}} + \frac{1}{2}\|f^\star\|_{\mathcal{H}_k}^2.$$

Taking expectation under $\mathbb{P}_{f^\star}$ proves the final claim. □

Combining Theorem 3.2 with Lemma 3.3, we obtain the following corollary, which gives a MAIR regret bound for Gaussian-process bandits. This is a central consequence of the framework: the bound is stated for arbitrary GP-bandit algorithms through their associated algorithmic coefficient.

Theorem 3.4 (MAIR regret bound for Gaussian-process bandits)**.** *Under the setup of Section 2 and Theorem 3.2, let $q_1 = \mathsf{GP}(0, k)$ be the Gaussian-process reference prior, and let q_t denote the corresponding posterior after $t-1$ observations. Then*

$$R_T \le \mathbb{E}_{f^\star}\left[\mathcal{I}_T^{\mathrm{ker}}\right] + \frac{1}{2}\|f^\star\|_{\mathcal{H}_k}^2 + \mathbb{E}\left[\sum_{t=1}^{T}\left\{\mathsf{MAIR}_{\rho_t,\eta}(p_t,\mu_t) + \left\langle \frac{\partial}{\partial\mu}\mathsf{MAIR}_{\rho_t,\eta}(p_t,\mu)\Big|_{\mu=\mu_t}, \mathbf{1}(M^\star) - \mu_t \right\rangle\right\}\right]$$

The following lemma is the closed-form gradient calculation behind the post-validation theorem. It is a continuous analogue of the finite calculation in Appendix 3.5 of Xu and Zeevi (2025).

Lemma 3.5 (MAIR gradient bracket)**.** *Assume the map $M \mapsto P_{M,\pi}$ is dominated and all displayed derivatives exist. Fix p, ρ, μ. Up to an additive constant on the probability simplex,*

$$\frac{\partial}{\partial\mu(M)}\mathsf{MAIR}_{\rho,\eta}(p,\mu) = \Delta_M(p) - \frac{1}{\eta}\log\frac{\mu(M)}{\rho(M)} - \frac{1}{\eta}\mathbb{E}_{\pi\sim p}\,\mathrm{KL}(P_{M,\pi}\|\mu_{o|\pi}).$$

Consequently, for any true model $M^\star$,

$$\begin{aligned}&\mathsf{MAIR}_{\rho,\eta}(p,\mu) + \langle\nabla_\mu\mathsf{MAIR}_{\rho,\eta}(p,\mu),\, \delta_{M^\star} - \mu\rangle\\ &\quad= \Delta_{M^\star}(p) - \frac{1}{\eta}\mathbb{E}_{\pi\sim p}\,\mathrm{KL}(P_{M^\star,\pi}\|\mu_{o|\pi}) - \frac{1}{\eta}\log\frac{\mu(M^\star)}{\rho(M^\star)}.\end{aligned}$$

In particular, when $\mu = \rho$,

$$\mathsf{MAIR}_{\mu,\eta}(p,\mu) + \langle\nabla_\mu\mathsf{MAIR}_{\mu,\eta}(p,\mu),\, \delta_{M^\star} - \mu\rangle = \Delta_{M^\star}(p) - \frac{1}{\eta}\mathbb{E}_{\pi\sim p}\,\mathrm{KL}(P_{M^\star,\pi}\|\mu_{o|\pi}). \tag{2}$$

Proof. Write

$$I_\mu(M; O \mid \pi) = \int \mu(dM)\int \log\frac{dP_{M,\pi}}{d\mu_{o|\pi}}\,dP_{M,\pi}.$$

The Gateaux derivative of $\mathbb{E}_\mu\Delta_M(p)$ in direction δ_M is $\Delta_M(p)$. The derivative of $\mathrm{KL}(\mu\|\rho)$ is $\log(\mu(M)/\rho(M))+1$, and the $+1$ is a simplex constant. The derivative of the mutual-information term is

$$\mathrm{KL}(P_{M,\pi}\|\mu_{o|\pi})$$

up to the same simplex constant. This follows by differentiating the mixture predictive density: the derivative of $-\int\log d\mu_{o|\pi}\,d\mu_{o|\pi}$ contributes a constant after integration over the simplex. Averaging over $\pi\sim p$ gives the displayed gradient.

Pairing the gradient with $\delta_{M^\star} - \mu$ gives

$$\begin{aligned}&\Delta_{M^\star}(p) - \mathbb{E}_\mu\Delta_M(p) - \frac{1}{\eta}\log\frac{\mu(M^\star)}{\rho(M^\star)} + \frac{1}{\eta}\mathrm{KL}(\mu\|\rho)\\ &\quad - \frac{1}{\eta}\mathbb{E}_{\pi\sim p}\,\mathrm{KL}(P_{M^\star,\pi}\|\mu_{o|\pi}) + \frac{1}{\eta}\mathbb{E}_{\pi\sim p}I_\mu(M; O \mid \pi).\end{aligned}$$

Adding $\mathsf{MAIR}_{\rho,\eta}(p,\mu)$ cancels the $\mathbb{E}_\mu\Delta$, mutual-information, and $\mathrm{KL}(\mu\|\rho)$ terms, yielding the claim. The case $\mu = \rho$ removes the model-prior log term. □

Equation (2) is the key post-validation identity. It says that the MAIR objective plus the belief-gradient error equals the true-model regret minus a true-model information penalty. It is exact. It does not yet bound regret; the regret term must still be controlled. For GP-UCB, that control comes from self-normalized calibration and optimism.

4 Frequentist GP-UCB as MAIR post-validation

We now apply the posterior-reference form above to GP-UCB. The action law is random through the history, even though it is conditionally degenerate,

$$p_t(\cdot \mid H_{t-1}) = \delta_{x_t(H_{t-1})},$$

and the reference belief ρ_{t-1} is the current surrogate GP posterior over $\widehat{f}$. The fixed function $f^\star$ is not assumed to be sampled from this posterior. The role of ρ_{t-1} is to define the posterior mean, variance, and one-step algorithmic information $g_t^{\mathrm{MAIR}} = I_{\rho_{t-1}}(\widehat{f}; Y_t \mid H_{t-1}, x_t)$; the frequentist verification is the calibration-plus-optimism argument below.

4.1 Self-normalized calibration

The calibration theorem is the frequentist bridge from the fake GP posterior to the fixed RKHS function.

Theorem 4.1 (Realized-information RKHS calibration). *With probability at least $1-\delta$, simultaneously for all $t \geq 0$ and $x \in \mathcal{X}$,*

$$|f^\star(x) - \mu_t(x)| \leq \beta_t^{\mathrm{RI}} \sigma_t(x),$$

where one may take, up to the standard time-uniform logarithmic adjustment,

$$\beta_t^{\mathrm{RI}} = B + \frac{R}{\sqrt{\lambda}} \sqrt{2\left(\mathcal{I}_t^{\mathrm{ker}} + \log\frac{1}{\delta}\right)}.$$

Proof. The KRR error decomposes as

$$f^\star(x) - \mu_t(x) = \lambda\langle f^\star, V_t^{-1}\phi(x)\rangle - \langle S_t, V_t^{-1}\phi(x)\rangle.$$

The deterministic bias term obeys

$$|\lambda\langle f^\star, V_t^{-1}\phi(x)\rangle| \leq \sqrt{\lambda}\|f^\star\|_{\mathcal{H}_k}\sqrt{\langle \phi(x), V_t^{-1}\phi(x)\rangle} = B\sigma_t(x).$$

The noise term satisfies

$$|\langle S_t, V_t^{-1}\phi(x)\rangle| \leq \|S_t\|_{V_t^{-1}}\|\phi(x)\|_{V_t^{-1}} = \frac{\sigma_t(x)}{\sqrt{\lambda}}\|S_t\|_{V_t^{-1}}.$$

The RKHS self-normalized martingale inequality gives, with probability at least $1-\delta$,

$$\|S_t\|_{V_t^{-1}} \leq R\sqrt{2\left(\frac{1}{2}\log\det(I+\lambda^{-1}K_t) + \log\frac{1}{\delta}\right)} = R\sqrt{2\left(\mathcal{I}_t^{\mathrm{ker}} + \log\frac{1}{\delta}\right)}$$

simultaneously over t; a peeling or mixture argument gives the fully time-uniform version. Combining the two inequalities proves the claim. □

This theorem is exactly where the hard frequentist work lives. MAIR does not make the surrogate prior true. It organizes the proof once calibration is available.

4.2 The MAIR error condition for GP-UCB

Run GP-UCB with realized-information coefficient

$$x_t \in \operatorname*{argmax}_{x\in\mathcal{X}} \left\{\mu_{t-1}(x) + \beta^{\mathrm{RI}}_{t-1}\sigma_{t-1}(x)\right\}.$$

Let $x^\star \in \operatorname{argmax}_x f^\star(x)$ and $r_t = f^\star(x^\star) - f^\star(x_t)$.

Lemma 4.2 (Optimism verifies the MAIR error condition)**.** *On the calibration event of Theorem 4.1, for every t,*

$$r_t \le 2\beta^{\mathrm{RI}}_{t-1}\sigma_{t-1}(x_t).$$

Let

$$g^{\mathrm{MAIR}}_t := I_{\rho_{t-1}}(\widehat{f}; Y_t \mid H_{t-1}, x_t) = \frac{1}{2}\log\left(1 + \lambda^{-1}\sigma^2_{t-1}(x_t)\right).$$

Then

$$r_t^2 \le 4C_{\lambda,\kappa}(\beta^{\mathrm{RI}}_{t-1})^2 g^{\mathrm{MAIR}}_t,$$

where

$$C_{\lambda,\kappa} := \frac{2\kappa^2}{\log(1+\kappa^2/\lambda)}.$$

Equivalently, for every $\eta > 0$,

$$r_t - \frac{1}{\eta}g^{\mathrm{MAIR}}_t \le \eta C_{\lambda,\kappa}(\beta^{\mathrm{RI}}_{t-1})^2. \tag{3}$$

Proof. On the calibration event,

$$f^\star(x^\star) \le \mu_{t-1}(x^\star) + \beta^{\mathrm{RI}}_{t-1}\sigma_{t-1}(x^\star).$$

By the definition of x_t,

$$\mu_{t-1}(x^\star) + \beta^{\mathrm{RI}}_{t-1}\sigma_{t-1}(x^\star) \le \mu_{t-1}(x_t) + \beta^{\mathrm{RI}}_{t-1}\sigma_{t-1}(x_t).$$

Again by calibration,

$$\mu_{t-1}(x_t) \le f^\star(x_t) + \beta^{\mathrm{RI}}_{t-1}\sigma_{t-1}(x_t).$$

Combining gives $r_t \le 2\beta^{\mathrm{RI}}_{t-1}\sigma_{t-1}(x_t)$.

For $0 \le z \le \kappa^2/\lambda$, concavity of $\log(1+z)$ gives

$$\lambda z = \sigma^2 \le C_{\lambda,\kappa}\frac{1}{2}\log(1+z).$$

Substituting $z = \lambda^{-1}\sigma^2_{t-1}(x_t)$ gives

$$\sigma^2_{t-1}(x_t) \le C_{\lambda,\kappa} g^{\mathrm{MAIR}}_t.$$

Squaring the optimism bound proves the ratio inequality. The offset inequality is Young's inequality $2\sqrt{ab} \le \eta a + b/\eta$ with $a = C_{\lambda,\kappa}(\beta^{\mathrm{RI}}_{t-1})^2$ and $b = g^{\mathrm{MAIR}}_t$. □

Combining the posterior-reference bracket with Lemma 4.2, define

$$\mathfrak{B}_t := \mathsf{MAIR}_{\rho_{t-1},\eta}(p_t, \rho_{t-1}) + \left\langle \nabla_\mu \mathsf{MAIR}_{\rho_{t-1},\eta}(p_t,\mu)\big|_{\mu=\rho_{t-1}}, \delta_{f^\star} - \rho_{t-1}\right\rangle.$$

By Lemma 3.5, with $p_t = \delta_{x_t}$,

$$\mathfrak{B}_t = r_t - \frac{1}{\eta}\mathrm{KL}(P_{f^\star,x_t}\|\rho_{t-1,o|x_t}) \le r_t.$$

The posterior-reference prior therefore leaves only the realized algorithmic information increment

$$g_t^{\mathrm{MAIR}} = \frac{1}{2}\log\left(1+\lambda^{-1}\sigma_{t-1}^2(x_t)\right),$$

and Lemma 4.2 gives

$$\mathfrak{B}_t \le \frac{1}{\eta} g_t^{\mathrm{MAIR}} + \eta C_{\lambda,\kappa}(\beta_{t-1}^{\mathrm{RI}})^2.$$

This is the GP-UCB MAIR error condition: the gradient theorem gives the bracket, while calibration and optimism bound it for the fixed $f^\star$.

Theorem 4.3 (GP-UCB through MAIR post-validation)**.** *On the calibration event of Theorem 4.1, GP-UCB satisfies*

$$R_T \le 2\sqrt{C_{\lambda,\kappa}\mathcal{I}_T^{\mathrm{ker}}\sum_{t=1}^{T}(\beta_{t-1}^{\mathrm{RI}})^2}.$$

Consequently,

$$R_T \le 2\beta_T^{\mathrm{RI}}\sqrt{C_{\lambda,\kappa}T\mathcal{I}_T^{\mathrm{ker}}}.$$

In particular,

$$R_T \lesssim \left[B + \frac{R}{\sqrt{\lambda}}\sqrt{\mathcal{I}_T^{\mathrm{ker}} + \log(1/\delta)}\right]\sqrt{T\mathcal{I}_T^{\mathrm{ker}}}.$$

Using $\mathcal{I}_T^{\mathrm{ker}} \le \gamma_T$ recovers the usual maximal-information-gain form.

Proof. Apply Corollary 3.4 (MAIR regret bound for Gaussian-process bandits) with

$$g_t = g_t^{\mathrm{MAIR}} = \frac{1}{2}\log\left(1+\lambda^{-1}\sigma_{t-1}^2(x_t)\right).$$

The finite $\log|\mathcal{M}|/\eta$ prior-mass potential is replaced by the realized budget $\sum_{t\le T} g_t/\eta = \mathcal{I}_T^{\mathrm{ker}}/\eta$. Lemma 4.2 gives the corresponding one-step coefficient certificate.

The realized information-ratio coefficient satisfies

$$\Psi_t^{\text{GP-UCB}} := \frac{r_t^2}{g_t^{\mathrm{MAIR}}} \le 4C_{\lambda,\kappa}(\beta_{t-1}^{\mathrm{RI}})^2.$$

Hence by Cauchy–Schwarz,

$$R_T = \sum_{t=1}^{T} r_t \le \sqrt{\left(\sum_{t=1}^{T} g_t^{\mathrm{MAIR}}\right)\left(\sum_{t=1}^{T}\Psi_t^{\text{GP-UCB}}\right)}.$$

Since $\sum_t g_t^{\mathrm{MAIR}} = \mathcal{I}_T^{\mathrm{ker}}$, the first display follows. Monotonicity of β_t^{RI} gives the second display, and substituting the calibration coefficient gives the third. □

Remark 4.4 (What this proves)**.** *Theorem 4.3 is not a new minimax theorem. It is a rigorous frequentist MAIR proof of the known GP-UCB bound in realized-information form. It clarifies that γ_T is a worst-case replacement for $\mathcal{I}_T^{\mathrm{ker}}$, while the genuinely difficult part of the COLT open problem is the coefficient $(\beta_t^{\mathrm{RI}})^2 \simeq \mathcal{I}_t$. A further improvement must replace this coefficient by a smaller calibrated AIR coefficient.*

5 Effective-ratio AIR coefficients

The information-ratio method historically bounds regret by relating one-step regret to one-step information. For finite-dimensional linear bandits, Russo and Van Roy (2014) control the information ratio by rank. The same argument has a kernel/operator version.

At a fixed round, let D be a compact operator or matrix of predictive mean shifts. Think of rows as possible optimal labels and columns as sampled decisions, weighted by the algorithmic belief and the decision law. Suppose

$$\Delta \leq L\|D\|_*, \qquad g \geq c\|D\|_F^2 \tag{4}$$

for constants $L, c > 0$.

Theorem 5.1 (Effective-ratio AIR bound)**.** *Under* (4),

$$\Psi = \frac{\Delta^2}{g} \leq \frac{L^2}{c} \frac{\|D\|_*^2}{\|D\|_F^2}.$$

Moreover, if D_r is the rank-r truncation of D and $\tau_r(D) := \|D - D_r\|_*$, then

$$\Psi \leq \frac{L^2}{c} \inf_{r\geq 1} \left(\sqrt{r} + \frac{\tau_r(D)}{\|D\|_F}\right)^2.$$

Proof. The first claim follows immediately from (4). For the approximation-aware claim, write

$$\|D\|_* \leq \|D_r\|_* + \|D - D_r\|_* \leq \sqrt{r}\|D_r\|_F + \tau_r(D) \leq \sqrt{r}\|D\|_F + \tau_r(D).$$

Substituting this into the first display and optimizing over r proves the theorem. □

Define

$$\mathrm{er}(D) = \frac{\|D\|_*^2}{\|D\|_F^2}, \qquad \mathrm{er}_{\mathrm{app}}(D) = \inf_{r\geq 1} \left(\sqrt{r} + \frac{\tau_r(D)}{\|D\|_F}\right)^2.$$

The quantity $\mathrm{er}_{\mathrm{app}}(D)$ is the effective AIR coefficient. It is a coefficient replacement for the confidence-width use of γ_t, not a replacement for cumulative information. It can be computed from a finite dictionary by singular-value decomposition. In large kernel problems, D can be approximated by Nyström inducing points, random Fourier features, or a regularized nearest-neighbor projection. The approximation error $\tau_r(D)$ is the price of using a smaller computational representation.

Comparison to maximal information gain. Maximal information gain is a cumulative scalar: it upper-bounds $\sum_t g_t$. The effective ratio is a one-step coefficient: it upper-bounds Δ_t^2/g_t. The classical GP-UCB analysis effectively uses

$$\Psi_t \lesssim (\beta_t^{\mathrm{RI}})^2 \simeq \mathcal{I}_t,$$

which gives $\mathcal{I}_T\sqrt{T}$ after summation. A successful effective-ratio analysis would prove

$$\Psi_t \lesssim \mathrm{er}_{\mathrm{app}}(D_t) \ll \mathcal{I}_t$$

along the algorithm's trajectory, giving

$$R_T \lesssim \sqrt{\mathcal{I}_T \sum_{t\leq T} \mathrm{er}_{\mathrm{app}}(D_t)}.$$

This is the AIR form of the hoped-for improvement from $\gamma_T\sqrt{T}$ toward $\sqrt{dT\gamma_T}$.

Regularized nearest-neighbor implementation. Let $\mathcal{D}_t = \{z_1, \dots, z_m\}$ be a dictionary of candidate or inducing actions. Let Σ_t be the ridge design covariance in feature space and define the local leverage

$$\ell_t(u; \mathcal{D}_t) = \min_{w \in \mathbb{R}^m} \left\| \phi(u) - \sum_{j=1}^{m} w_j \phi(z_j) \right\|_{\Sigma_t + \lambda I}^2 + \alpha \|w\|_1.$$

A nearest-neighbor AIR algorithm chooses a decision distribution p_t or optimistic action by minimizing an offset objective with coefficient

$$\mathcal{K}_t^{\mathrm{NN}} = \sup_{u \in \mathcal{X}} \ell_t(u; \mathcal{D}_t)$$

or a sampled/weighted version of this supremum. This is not an oracle subatlas: the dictionary is generated from data, inducing points, or randomized features, and the coefficient is directly computable.

6 Hub–cloud separation for full-action regularized GP-UCB

This section presents a finite GP-bandit separation showing that, in overparameterized models, algorithmic complexity may provide a sharper certificate than full-class minimax complexity. For related background and motivation on the "hub–leaves–cloud" construction in the offline setting, see Xu (2026); Li and Xu (2026).

Let

$$\mathcal{X} = H_M \cup C_N, \qquad H_M = \{h_0, h_1, \dots, h_M\}, \qquad C_N = \{c_1, \dots, c_N\}.$$

Consider independent Gaussian observation noise with variance σ^2. The kernel is diagonal, $k(x, x') = \mathbf{1}\{x = x'\}$, so this is a finite GP bandit. The loss is usual cumulative regret without complexity penalty.

A complexity-weighted GP-UCB rule uses the full action set but an agnostic complexity prior

$$q_0(x) \propto \exp(-\tau c(x)), \qquad c(h) = 0 \ (h \in H_M), \quad c(c) = L \ (c \in C_N),$$

or equivalently the acquisition

$$x_t \in \operatorname*{argmax}_{x \in \mathcal{X}} \left\{ \mu_{t-1}(x) + \beta_t \sigma_{t-1}(x) - \tau c(x) \right\}. \tag{5}$$

The algorithm is not restricted to the hub; it sees all $M + N + 1$ arms. The prior is algorithmic. It expresses that high-complexity cloud arms require evidence before being chosen.

Theorem 6.1 (Hub-scale performance under hub truths)**.** *Let the true reward be a hub truth: one hub arm $h^\star$ has mean $R > 0$, all other arms have mean 0. Suppose the UCB calibration event holds for all arms up to horizon T and*

$$\tau L \geq 2\beta_T + R.$$

Then the full-action regularized GP-UCB rule (5) never selects a cloud arm. Consequently,

$$R_T \leq 2\beta_T \sum_{t=1}^{T} \sigma_{t-1}(x_t) \lesssim \beta_T \sqrt{(M+1) T \log(1 + T/\sigma^2)}.$$

Proof. For any cloud arm c, the confidence event and the fact that its true mean is 0 imply

$$\mu_{t-1}(c) + \beta_t \sigma_{t-1}(c) - \tau L \leq 2\beta_T - \tau L \leq -R.$$

For the optimal hub arm $h^\star$, calibration gives

$$\mu_{t-1}(h^\star) + \beta_t \sigma_{t-1}(h^\star) \geq f^\star(h^\star) = R.$$

Thus every cloud index is strictly below the index of $h^\star$, so no cloud arm is selected. The regret analysis then reduces to finite-arm UCB on $M+1$ independent arms. On the confidence event,

$$r_t \leq 2\beta_t \sigma_{t-1}(x_t),$$

and the standard variance-counting bound for independent arms gives

$$\sum_{t=1}^{T} \sigma_{t-1}(x_t) \lesssim \sqrt{(M+1)T \log(1 + T/\sigma^2)}.$$

□

Now consider the full ambient class containing cloud truths: for each $j \leq N$, model M_j has mean S on cloud arm c_j and mean 0 elsewhere. This class is part of the same finite GP/DMSO model; there is no representation ambiguity.

Proposition 6.2 (Full-class cloud lower bound)**.** *For any algorithm and any horizon $T \leq N/4$,*

$$\sup_{j \leq N} \mathbb{E}_{M_j} R_T \geq \frac{ST}{2}.$$

Moreover, for the one-step offset DEC with reference model M_0 having all means zero, there exists a universal constant $c > 0$ such that

$$\mathsf{DEC}_\gamma(\{M_0, M_1, \ldots, M_N\}, M_0) \geq \frac{S}{4} \quad \textit{whenever} \quad \gamma \leq c\frac{\sigma^2 N}{S}.$$

Proof. For the minimax lower bound, draw J uniformly from $\{1, \ldots, N\}$ and let τ_J be the first time the algorithm plays c_J. Before τ_J, all observations from cloud arms have the same distribution under every J not yet queried. Hence, conditional on no previous hit and on the set of previously queried cloud arms, J is uniform over the remaining cloud arms. Therefore

$$\mathbb{P}(\tau_J = t \mid \tau_J \geq t) \leq \frac{1}{N - t + 1},$$

and for $T \leq N/4$,

$$\mathbb{P}(\tau_J \leq T) \leq \sum_{t=1}^{T} \frac{1}{N - t + 1} \leq \frac{T}{N - T} \leq \frac{1}{3}.$$

On the event $\tau_J > T$, every selected arm has mean zero while the optimal cloud arm has mean S, so regret is ST. Thus the Bayes regret under the uniform prior is at least $2ST/3$, and in particular at least $ST/2$. Minimax regret is at least Bayes regret.

For the one-step DEC lower bound, let p be any distribution over arms. There exists j with $p(c_j) \leq 1/N$. Under M_j, the one-step regret is at least $S(1 - p(c_j)) \geq S(1 - 1/N)$. The observation law differs from the reference only when c_j is played. For Gaussian noise,

$$\mathrm{KL}(P^p_{M_j} \| P^p_{M_0}) = p(c_j)\frac{S^2}{2\sigma^2} \leq \frac{S^2}{2\sigma^2 N},$$

and squared Hellinger is bounded by KL. Therefore the offset regret-information expression is at least

$$S(1 - 1/N) - \gamma\frac{S^2}{2\sigma^2 N},$$

which is at least $S/4$ for $N \geq 2$ and $\gamma \leq c\sigma^2 N/S$ with c small enough. □

The gap between Theorem 6.1 and Proposition 6.2 can be arbitrarily large by taking $N \gg M$. It is not an estimation-complexity gap: the class is finite and observable. It is not a Bellman-rank vacuity: there is no hidden witness choice. It is an AIR/minimax gap inside a clean GP-bandit model. The regularized GP-UCB algorithm is practical precisely because it does not optimize for full-class robustness against high-complexity cloud truths.

7 Bellman rank, DEC, and the IR–DEC duality

This paper uses DMSO in the standard sense: a decision space, model class, observation laws, and rewards are all explicitly specified. DEC is powerful in this setting because it is a clean class-wide offset complexity. A phrase such as "bounded Bellman rank" is different. Bellman rank is a representation-relative algebraic certificate depending on value-function classes, roll-in distributions, and Bellman-error witnesses (Jiang et al., 2017). Without fixing the witness/discrepancy and the observable experiment, bounded Bellman rank is not itself a DMSO model class and cannot by itself be assigned an intrinsic DEC. Bellman rank says that there exists a low-dimensional witness structure under which a class is learnable. DEC says that a fixed observable model class has a particular decision-estimation tradeoff. Once a bilinear-class witness (Du et al., 2021) is fixed and its estimation complexity is charged, one can ask DEC-style questions. Without that, an arbitrary rich witness can make a lower bound vacuous or non-intrinsic.

There are also gaps within clean DMSO classes. Chen et al. (2024) show that DEC-style lower bounds need estimation-complexity refinements such as fractional covering to recover tight lower bounds in general interactive problems. Work on ridge-bandit and even MAB models (Rajaraman et al., 2023; Gu et al., 2025) show another lesson: minimax regret may not be recovered from a scalar one-round coefficient alone; algorithm dynamics and phases matter.

The hub–cloud separation here is different. It does not exploit representation ambiguity or an omitted estimation term. The full class is finite, the observations are Gaussian, and the regret loss is ordinary. The full-class minimax/DEC benchmark is large because the class contains a huge cloud of high-complexity alternatives. A regularized full-action GP-UCB algorithm intentionally refuses to be robust to those alternatives unless evidence accumulates. Its algorithmic AIR performance on hub truths is therefore much smaller than the full-class DEC certificate.

7.1 Historical and conceptual comparison

Information-ratio methods were born algorithmic: Russo and Van Roy (2016, 2014) introduced the ratio to analyze Thompson sampling and information-directed sampling. The mirror-descent route of Lattimore and György (2021) moved this idea from Bayesian to frequentist learning. Foster et al. (2021, 2022) proposed DEC as an intrinsic class-wide offset complexity for DMSO problems and they are especially powerful for lower bounds and minimax algorithms. We refer to Xu and Zeevi (2025) for further background and discussion. That work builds on both lines, and its exchange-of-decision duality route simplifies the nonconstructive functional-estimator step in Lattimore and György (2021).

Ignoring historical influence and focusing on mathematical form, IR/AIR and DEC are dual offset views. IR/MAIR conditions on the algorithmic belief and decision/model marginal, then measures how much posterior-update information the algorithm obtains from its own action. DEC fixes a clear model class and reference model, then asks for the best decision distribution against the worst model. IR/AIR is naturally suited to many concrete Bayesian/frequentist algorithms: TS/IDS, exponential-weights, inverse-propensity-weighted procedures, EBO, APS/AMS algorithms, and, as shown above, UCB families after calibration. DEC is naturally suited to minimax lower bounds, robust class-wide algorithms, and black-box estimators.

	Scalar / class-wide coefficient	Algorithmic / pointwise coefficient
IR / AIR / MAIR	Classical information ratio; model-index information ratio; often summarized by a single dimension or rank. Gives upper bounds for Bayesian or frequentist algorithms once calibrated.	Algorithm-specific offset: realized AIG plus AIR coefficient. Applies to actual policies, including TS, IDS, exponential weights, IPW-style rules, and GP-UCB after MAIR post-validation.
DEC / convex-hull, and model-free variants	Class-wide DMSO complexity. Excellent for minimax lower bounds and robust algorithms. Convex-hull and reference-model formulations make the model class explicit.	Localized or restricted DEC variants can certify subfamilies, but they answer a different question unless the restriction is itself part of the model/algorithm specification.

The conclusion is not that one framework dominates the other. DEC is more intrinsic to well-specified model classes. IR/AIR is more algorithmic by construction. In overparameterized GP/kernel bandits, practical algorithms often use algorithmic priors and regularization; their meaningful complexity is therefore the realized information and coefficient of the algorithm, not necessarily the full-class minimax certificate.

8 Conclusion

GP-UCB is a paradigmatic algorithmic-prior method. In frequentist RKHS theory, the surrogate GP prior is not true; it is a computational and inferential device. The classical self-normalized martingale theorem calibrates that device against a fixed RKHS function. Once calibrated, GP-UCB's optimism verifies the MAIR error condition and yields a realized-information regret proof.

This perspective gives a precise interpretation of maximal information gain. Maximal information gain is a useful worst-case benchmark, but it is not the intrinsic algorithmic object. The realized log determinant is the algorithmic information gain of the trajectory. The open confidence-width problem is about the coefficient: can the self-normalized $\mathcal{I}_t$-scale coefficient be replaced by a smaller calibrated AIR coefficient, such as a local effective ratio?

The hub–cloud example explains why this question matters. In overparameterized models, class-wide minimax robustness can be dominated by irrelevant high-complexity alternatives. Practical GP-UCB-style algorithms often succeed because their algorithmic priors suppress such alternatives. MAIR/AIR provides the right language for proving and comparing such algorithmic guarantees, while DEC remains the right language for well-specified class-wide minimax questions.

Acknowledgments

The manuscript is written to be self-contained. The MAIR post-validation proof of GP-UCB and the separation from full-class minimax complexity are rigorous. The effective-ratio coefficient theorem is an abstract coefficient bound; potential improvement over the classical GP-UCB rate would require either a new calibration argument or new control of the algorithmic coefficient. The author thanks Wenjia Wang for bringing this problem to his attention.

References

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.

Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *Proceedings of the 34th International Conference on Machine Learning*, 2017.

Simon S. Du, Sham M. Kakade, Jason D. Lee, Shachar Lovett, Gaurav Mahajan, Wen Sun, and Ruosong Wang. Bilinear classes: A structural framework for provable generalization in RL. In *Proceedings of the 38th International Conference on Machine Learning*, PMLR 139:2826–2836, 2021.

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, Alexander Rakhlin, Ayush Sekhari, and Karthik Sridharan. On the complexity of adversarial decision making. In *Advances in Neural Information Processing Systems 35*, 2022.

Yuzhou Gu, Yanjun Han, and Jian Qian. Evolution of information in interactive decision making: A case study for multi-armed bandits. *arXiv:2503.00273*, 2025.

Shogo Iwazaki. Improved regret bounds for Gaussian process upper confidence bound in Bayesian optimization. In *Advances in Neural Information Processing Systems 38*, 2025.

Nan Jiang, Akshay Krishnamurthy, Alekh Agarwal, John Langford, and Robert E. Schapire. Contextual decision processes with low Bellman rank are PAC-learnable. In *Proceedings of the 34th International Conference on Machine Learning*, 2017.

Tor Lattimore and András György. Mirror descent and the information ratio. In *Proceedings of the 34th Annual Conference on Learning Theory*, PMLR 134:2965–2992, 2021.

Tor Lattimore. A lower bound for adaptive linear regression. In *Proceedings of the 34th International Conference on Algorithmic Learning Theory*, 2023.

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

Nived Rajaraman, Yanjun Han, Jiantao Jiao, and Kannan Ramchandran. Statistical complexity and optimal algorithms for non-linear ridge bandits. *arXiv:2302.06025*, 2023.

Daniel Russo and Benjamin Van Roy. Learning to optimize via information-directed sampling. In *Advances in Neural Information Processing Systems 27*, 2014.

Daniel Russo and Benjamin Van Roy. An information-theoretic analysis of Thompson sampling. *Journal of Machine Learning Research*, 17:1–30, 2016.

Jonathan Scarlett, Ilija Bogunovic, and Volkan Cevher. Lower bounds on regret for noisy Gaussian process bandit optimization. In *Proceedings of the 30th Annual Conference on Learning Theory*, 2017.

Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proceedings of the 27th International Conference on Machine Learning*, 2010.

Shion Takeno and Shogo Iwazaki. On regret bounds of Thompson sampling for Bayesian optimization. *arXiv:2603.09276*, 2026.

Sattar Vakili, Jonathan Scarlett, and Tara Javidi. Open problem: Tight online confidence intervals for RKHS elements. In *Proceedings of the 34th Annual Conference on Learning Theory*, PMLR 134:4647–4652, 2021.

Sattar Vakili, Kia Khezeli, and Victor Picheny. On information gain and regret bounds in Gaussian process bandits. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, PMLR 130:82–90, 2021.

Wenjia Wang and Yifang Zhang. On the suboptimality of GP-UCB under polynomial effective optimism. *arXiv:2312.01386*, revised 2026.

Justin Whitehouse, Zhiwei Steven Wu, and Aaditya Ramdas. On the sublinear regret of GP-UCB. In *Advances in Neural Information Processing Systems 36*, 2023.

Yunbei Xu. Pointwise complexity for Gaussian fields: Upper envelopes, algorithmic lower bounds, and separation. Manuscript, 2026.

Yunbei Xu and Assaf Zeevi. Bayesian design principles for frequentist sequential learning. *Journal of the ACM*, 72(5):37, 2025. doi:10.1145/3766898.