BAYESIAN DECISION THEORY FOR GAUSSIAN PROCESS MODELS: WITH SOME RESULTS ON LOSS FUNCTION EQUIVALENCE AND A DECISION THEORETIC EXTENSION THEOREM

Ryan Warnick
Independent Work
rswarnick1@gmail.com

May 20, 2026

ABSTRACT

We provide a plug-and-drop decision theoretic recipe to recover source terms Q in general order linear ordinary differential equations ($\mathcal{L}y = \sum_{k=0}^{n} a_k(\circ)\, y^{(k)} = Q, \quad x \in \Omega$) and partial differential equations ($(\mathcal{L}y)(x) \;=\; \sum_{|\alpha|\leq k} A_\alpha(x)\, D^\alpha y(x), \quad x \in \Omega$), assuming a Gaussian process prior on y with derivatives up to order k in mean square. Under squared error loss, the Bayes action is the posterior mean of $\mathcal{L}y$; we give closed-form expressions for both the mean and covariance of Q, handle general linear observations (including exact/noisy boundary conditions), and provide an implementation-ready matrix formulation. The method permits evaluation of infinite-dimensional Bayes actions on function spaces (with associated uncertainty quantification), or, in the finite approximation setting, an uncountably infinite family of Bayes actions corresponding to the counting measure on the domain $x \in \Omega$. An equivalence between a type of 0-1 loss and the mean squared loss on Q is shown under Gaussian process measures, and an extension theorem for infinite-dimensional decision rules on function spaces is formalized. Additionally, an extension to nonlinear operators is also outlined utilizing second-order Frechet expansions, with associated probability bounds.

1 Introduction

The Bayesian decision-theoretic perspective offers a powerful and conceptually unified approach to inference, uncertainty quantification, and optimal decision making in parametric problems; however, extensions to the infinite-dimensional setting have been minimally explored. While Gaussian processes (GPs) have become a central tool for modeling and inference in statistical and machine learning applications, the decision-theoretic dimension of GP modeling concerned with the definition and computation of optimal actions under posterior uncertainty remains comparatively underexplored.

The following work establishes foundational results for Bayesian decision theory for Gaussian process models, proving the equivalence of 0–1 loss and mean-squared loss under Gaussian measures, and formalizing the extension theorem for Bayes actions on Hilbert spaces; an additional result on decision boundaries between two Gaussian processes have is outlined, allowing the design of Bayes-optimal decision boundaries between two distinct Gaussian processes defined on a shared Hilbert space. These results connect classical Bayesian inverse problems [Dashti and Stuart, 2015, Kaipio and Somersalo, 2004, Kaipio et al., 2007] to the emerging field of probabilistic numerics [Cockayne et al., 2019], where algorithms are treated as stochastic estimators and evaluated under decision theoretic loss functions.

In this continuation, we propose an extension the theoretical and practical scope of these proceeding frameworks by addressing three interconnected challenges: *(i)* The formal analysis of decision theoretic risk in infinite dimensional Hilbert space problems. *(ii)* The integration of decision-theoretic risk into experimental design and data acquisition. *(iii)* The extension of risk bounds and uncertainty quantification to nonlinear and operator-valued mappings. Much of

the following paper is expository results outlining known properties of inverse problems in differential equations (both partial and ordinary) using GPs. However, the decision theoretic framing is novel, and the ensuing decision theoretic results outlined in Section 2 permit generalizations of the same approaches to broader classes of decision problems.

The paper then operates as a survey outlining general background and a computational approaches for source function evaluation from Gaussian processes (GPs) believed to be noisy realizations from an ODE or PDE, but also with extensions to nonlinear operators and three principal results extending GP based decision theory as well as decision theory on general Hilbert spaces.

1.1 Motivation

Many scientific and engineering problems involve making optimal decisions under uncertainty over continuous domains. Inverse problems, operator identification, and physical system modeling all depend on recovering unknown functions or operators from noisy data. Gaussian processes provide a nonparametric and flexible prior over such functions, while Bayesian inference supplies the engine for updating beliefs in light of data. Typically these are treated as separate problems, as the camp of practitioners in the scientific computing and inverse problems camp use the Bayesian machinery (and primarily GP-based machinery) only for functional inference, and then the ensuing inversion employs general tools from probabilistic-numerics [Cockayne et al., 2019, Dashti and Stuart, 2015]. However, in the Bayesian community more broadly there is broad effort to optimize decision making in uncertain environments through loss function design and finding optimal bayes actions through minimizing the posterior expected loss [Berger, 2010].

This combined perspective permits a decision theoretic and Bayesian experimental design [Chaloner and Verdinelli, 1995] approach to optimal decision making in Bayesian inverse problems (ordinary/partial differential equation based or otherwise). In the Gaussian process solution to inverse problems, this permits an additional benefit: the results here make no effort to diminutize the longstanding $O(n^3)$ for number of samples $n \in \mathbb{N}$ computational bottleneck of Gaussian process inference, however, if one can get $m \in \mathbb{N} < n$ for which the Bayes-risk under the m is less than under n, or which has lower variance of the Bayes action solution that minimizes Bayes-risk, then the computational bottleneck can be ever-so-slightly circumvented.

1.2 Contributions

This work makes the following contributions, expanding the theoretical and computational space of Bayesian decision theory for nonparametric models (and particularly GPs):

1. **We prove three principal novel results in Bayesian decision theory on Hilbert spaces**:
   - *Extension theorem for Bayes actions on Hilbert spaces* We prove that under mild regularity conditions on the loss function and underlying Hilbert space upwardly convergent subspaces (e.g. finite approximations to the underlying Hilbert space) converge to the solution to the loss acting on the global Hilbert space.
   - *0–1 small ball probability loss vs. squared-error loss Bayes action equivalence for Gaussian Process Measures on Hilbert spaces* An equivalence theorem between the squared error loss and small ball probability loss (a variant of 0-1 loss) is shown for Gaussian process measures under certain conditions. This extends the result outlining equivalence between the mean minimizing the L_2 loss in finite dimensional Gaussians and the median (which is the mean) minimizing the 0-1 loss in finite dimensional Gaussians.
   - *Decision boundary for two Gaussian-process posteriors* For two Gaussian processes defined on a shared Hilber Space, we give an explicit characterization of a Hilbert manifold contained in the said Hilbert space which is locally diffeomorphic to the global Hilbert manifold. E.g., the decision boundary between two GPs locally appears like the shared Hilbert space and Frechet calculus can be performed on it.
2. **Bayesian Risk-Based Experimental Design.** As outlined in section 1.1, we introduce a new formalism for optimal data acquisition based on posterior expected loss rather than posterior variance. This leads to a family of risk driven experimental design procedures that directly target decision functionals (e.g., differential operators, source terms, or physical observables) and can be efficiently computed from the GP posterior mean and covariance. This contribution bridges the gap between Gaussian process regression and Bayesian optimal design.
3. **Frechet-Based Nonlinear Operator Analysis.** We extend the Frechet derivative framework to quantify risk and uncertainty in nonlinear operator mappings applied to Gaussian process posteriors over function. The analysis establishes explicit high probability bounds on decision error as a function of the operator norm of the Frechet derivative and curvature. These results generalize the previous linear Gaussian guarantees outlined in the preceeding sections to the nonlinear setting and provide the theoretical underpinning for risk certification in operator learning and physics-informed modeling.

4. **Unified Computational Framework.** We provide a matrix-based formulation for implementing decision-theoretic inference and experimental design, compatible with general linear observation operators and flexible boundary conditions. The proposed algorithms admit efficient implementations via standard Gaussian process conditioning, while preserving theoretical rigor in infinite-dimensional Hilbert spaces.
5. **Applications and Implications.** We discuss applications of the developed theory to Bayesian inverse problems and probabilistic numerics methods, highlighting how decision theoretic risk can serve as a universal design and validation metric in Gaussian process modeling. This framework lays the groundwork for principled risk-aware learning in physics informed neural and kernel-based models.

1.3 Relation to Prior Work and Looking Forward

The framework developed here connects three strands of research. First, it builds on the theory of Bayesian inverse problems Dashti and Stuart [2015], Kaipio and Somersalo [2004], extending classical results on posterior consistency to decision-based optimality. Second, it aligns with probabilistic numerical methods Cockayne et al. [2019], where the objective is to quantify uncertainty in numerical computations. Finally, it contributes to the growing literature on physics-informed and operator-based learning Raissi et al. [2019], offering a theoretical foundation for uncertainty-aware operator inference. The framework similarly extends the rich Bayesian tradition within decision theory [Berger, 2010] to broader classes of decision problems on Hilbert spaces, and particularly, as outlined in the following Section, Theorem 1 serves as an extension theorem for any decision problem on Hilbert spaces for which the loss satisfies certain regularity conditions. This opens the door to other rich classes off nonparametric posteriors in Bayesian analysis.

1.4 Remaining Gap: Computational Bottleneck

Reiterating what was outlined in section 1.1, the most substantial missing component of the framework is that we do not address the substantial challenge of the $O(n^3)$ computational bottleneck of Gaussian process inference. However, a major contribution is that, as noted in Section 1.2, the Bayesian experimental design approach allows optimizing performance for computable n by optimizing the experimental design. This permits the practicioner to optimize for various types of loss functions and assess the overall contribution to reducing uncertainty in Source function estimation of each data point (and, more broadly, general Gaussian process decision problems), and reduce the posterior uncertainty optimally.

As outlined in Section 1.3, there exist many rich classes of Bayesian nonparametric priors and ensuing posterior formulations for which Theorem 1 permits analysis. Some of these include Polya-trees [Lavine, 1992], Gibbs-type species sampling priors [De Blasi et al., 2015], and completely random measures governed by a Levy-intensity [Kingman, 1967]. We hope this explicit foray into what at first pass appears to be the most tractable of all nonparametric priors (GPs) inspires others to pursue similar directions with broader families.

1.5 Paper Organization

The paper is organized as follows: Section 1 introduces the general problem definition, with 1.1 outlining the motivations for this work. Section 1.2 outlines the overall contributions presented here, first outlining notable theoretical contributions (with proofs contained in the Appendix) and then outlining contributions methodologically, including contributions to experimental design with GP priors, and extensions to nonlinear operators using Frechet bounds. The following two sub-sub-sections correspond to easy to use drop-in methods for incorporating Gaussian Processes into inverse problems in differential equations. Finally a section on applications and implications of the article. Following these sub-sub-sections is a subsection on Prior work and looking forward containted in subsection 1.3

2 Results

We outline three principal results and prove them in the appendix. These results support the decision theoretic framing and extend the mean squared loss as equivalent to a larger family of losses and decision problems.

Theorem 1 (Extension theorem for Bayes actions on Hilbert spaces)**.** *Let $\mathcal{H}$ be a real, separable Hilbert space. Let $\Psi : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ be a loss function, and let Q be a random element in $\mathcal{H}$ drawn from a posterior measure given data z. Let $\{H_n\}_{n=1}^{\infty}$ be an increasing sequence of finite-dimensional subspaces such that $\overline{\bigcup H_n} = \mathcal{H}$, and let $P_n : \mathcal{H} \to H_n$ be the corresponding orthogonal projection operators.*

Define the global expected posterior risk $R(a) = \mathbb{E}_{Q|z}[\Psi(a, Q)]$ and the finite-dimensional projected risk $R_n(a) = \mathbb{E}_{Q|z}[\Psi(a, P_n Q)]$. Assume the following conditions hold:

1. ***Growth Bound:*** *There exist $C > 0$ and $p \geq 1$ such that $|\Psi(a, h)| \leq C(1 + \|a\|^p + \|h\|^p)$, and $\mathbb{E}_{Q|z}[\|Q\|^p] < \infty$.*
2. ***Lipschitz Continuity in the State:*** *There exists $L > 0$ such that for all $a, h, h' \in \mathcal{H}$, $|\Psi(a, h) - \Psi(a, h')| \leq L\|h - h'\|$.*
3. ***Continuity in the Action:*** *For any fixed $h \in \mathcal{H}$, the map $a \mapsto \Psi(a, h)$ is continuous.*
4. ***Coercivity:*** *$R(a) \to \infty$ as $\|a\| \to \infty$.*
5. ***Uniform Convexity:*** *$R(a)$ is uniformly convex on $\mathcal{H}$.*

Then, for each n, there exists a unique minimizer $a_n = \arg\min_{a \in H_n} R_n(a)$. Furthermore, the sequence (a_n) converges strongly to the unique global Bayes action $a^* = \arg\min_{a \in \mathcal{H}} R(a)$; that is, $\lim_{n\to\infty} \|a_n - a^*\| = 0$.

Proof. Proof available in Appendix section A □

Theorem 2 (0–1 small ball probability loss vs. squared-error loss Bayes action equivalence for Gaussian Process Measures on Hilbert spaces)**.** *Let $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ be a real, separable Hilbert space equipped with the norm $\|\cdot\|$. Let Q be a random element in $\mathcal{H}$ whose posterior distribution given data z is Gaussian:*

$$Q \mid z \sim \mathcal{N}(\mu, \mathcal{C}) \tag{1}$$

where $\mu \in \mathcal{H}$ is the posterior mean, and $\mathcal{C} : \mathcal{H} \to \mathcal{H}$ is a self-adjoint, positive, trace-class covariance operator.

Let $\{H_n\}_{n=1}^{\infty}$ be an increasing sequence of finite-dimensional closed subspaces such that $H_n \uparrow \mathcal{H}$, and let $P_n : \mathcal{H} \to H_n$ denote the orthogonal projection operator onto H_n. For a localized action $a_n \in H_n$ and a fixed radius $\tau > 0$, define the sequence of finite-dimensional projected risk minimization problems under the L_2 loss and the $0-1$ small-ball loss respectively:

$$a^*_{2,n} = \arg\min_{a_n \in H_n} \mathbb{E}_{Q|z}\left[\|P_nQ - a_n\|^2 \mid z\right] \tag{2}$$

$$a^*_{01,n} = \arg\min_{a_n \in H_n} \mathbb{E}_{Q|z}\left[\mathbb{1}\{\|P_nQ - a_n\| > \tau\} \mid z\right] \tag{3}$$

Then, for every $n \geq 1$, the following properties hold:

1. ***Subspace Equivalence:*** *The unique minimizer on every finite-dimensional subspace under both losses is exactly the projected posterior mean:*
$$a^*_{2,n} = a^*_{01,n} = P_n\mu \tag{4}$$
2. ***Infinite-Dimensional Limit:*** *As $n \to \infty$, the sequence of subspace optimal actions converges strongly in $\mathcal{H}$ to the unique global infinite-dimensional Bayes action $a^* = \mu$:*
$$\lim_{n\to\infty} \|a^*_{2,n} - \mu\| = 0 \quad \textit{and} \quad \lim_{n\to\infty} \|a^*_{01,n} - \mu\| = 0 \tag{5}$$

Proof. Proof available in Appendix section B □

Theorem 3 (Decision boundary for two Gaussian-process posteriors)**.** *Let $\mathcal{H}$ be a real, separable Hilbert space. Let $Y \sim \mathcal{N}(\mu, \mathcal{C})$ and $Y' \sim \mathcal{N}(\mu', \mathcal{C}')$ be Gaussian random elements in $\mathcal{H}$ with means $\mu, \mu' \in \mathcal{H}$ and covariance operators $\mathcal{C}, \mathcal{C}'$ (self-adjoint, positive, trace-class on the closed span of their supports). For any $a \in \mathcal{H}$, define the small-ball* 0–1 *risks*

$$R^Y_{01,\tau}(a) := \mathbb{E}[\mathbf{1}\{\|Y - a\| > \tau\}], \qquad R^{Y'}_{01,\tau}(a) := \mathbb{E}[\mathbf{1}\{\|Y' - a\| > \tau\}].$$

Define the decision boundary *as the zero-level set of the limiting risk difference,*

$$\mathcal{B} := \Big\{a \in \mathcal{H} : \lim_{\tau\downarrow 0}\big(R^Y_{01,\tau}(a) - R^{Y'}_{01,\tau}(a)\big) = 0\Big\}.$$

Then $\mathcal{B}$ is a Hilbert manifold in $\mathcal{H}$, and coincides with the zero-level set of the squared-loss risk difference,

$$\mathcal{B} = \Big\{a \in \mathcal{H} : \mathbb{E}\|Y - a\|^2 = \mathbb{E}\|Y' - a\|^2\Big\},$$

and therefore admits the explicit affine description

$$\boxed{\|a - \mu\|^2 - \|a - \mu'\|^2 = \mathrm{tr}(\mathcal{C}') - \mathrm{tr}(\mathcal{C}).}$$

Which is an affine Hilbert manifold in $\mathcal{H}$ locally diffeomorphic to $\mathcal{H}$. In particular, when $\mathcal{C} = \mathcal{C}'$, the boundary is the (codimension-one) bisector hyperplane Hilbert manifold in $\mathcal{H}$

$$\{\, a \in \mathcal{H} : \langle a, \mu' - \mu \rangle = \tfrac{1}{2}(\|\mu'\|^2 - \|\mu\|^2) \,\},$$

i.e. the locus of points equidistant (in $\|\cdot\|$) from μ and μ'.

Proof. Proof available in Appendix Section C □

3 Linear k'th Order ODEs

3.1 Model and observations

Operator. Let

$$(\mathcal{L}y)(u) \;=\; \sum_{k=0}^{n} a_k(u)\, y^{(k)}(u), \quad u \in \mathcal{U} \subset \mathbb{R},$$

with known coefficient functions $a_k : \mathcal{U} \to \mathbb{R}$ and order $n \geq 0$.

Prior. Place a zero-mean GP prior on the latent solution:

$$y \sim \mathrm{GP}\big(0, K\big),$$

with kernel K differentiable up to order n in each argument.

Observations (very general). Assume we observe $\mathbf{z} \in \mathbb{R}^m$ through a *linear* observation operator $\mathcal{H}$ with additive Gaussian noise:

$$\mathbf{z} \;=\; \mathcal{H}[y] \;+\; \varepsilon, \qquad \varepsilon \sim \mathrm{N}(0, R).$$

This covers point evaluations $y(t_i)$, derivative samples $y^{(j)}(t_i)$, linear functionals (e.g., integrals), and (noiseless) boundary constraints by taking the corresponding rows of R to have vanishing variance (practically, a very small jitter).

3.2 Decision problem and Bayes estimator

We seek a decision rule $\hat{Q} : \mathcal{U} \to \mathbb{R}$ minimizing posterior expected squared loss

$$\mathcal{R}(\hat{Q}) \;=\; \mathbb{E}\Big[\int_{\mathcal{U}} \big(\mathcal{L}y(u) - \hat{Q}(u)\big)^2 \,\mathrm{d}u \,\Big|\, \mathbf{z}\Big].$$

By convexity and linearity of expectation, the pointwise minimizer is

$$\boxed{\hat{Q}(u) \;=\; \mathbb{E}[(\mathcal{L}y)(u) \,|\, \mathbf{z}] \;=\; \sum_{k=0}^{n} a_k(u)\, \mathbb{E}\Big[y^{(k)}(u) \,\Big|\, \mathbf{z}\Big].}$$

Uncertainty (useful for UQ and design) comes from the posterior covariance

$$\mathrm{Cov}(Q(u), Q(v) \mid \mathbf{z}) \;=\; \mathrm{Cov}((\mathcal{L}y)(u), (\mathcal{L}y)(v) \mid \mathbf{z}).$$

3.3 Closed forms via linear-Gaussian conditioning

Let $U_* = \{u_1, \ldots, u_s\}$ be test locations where we want Q. Stack the derivatives of y needed at U_* as

$$\mathbf{y}_* \;=\; \big[\, y^{(0)}(U_*);\; y^{(1)}(U_*);\; \ldots;\; y^{(n)}(U_*) \,\big] \in \mathbb{R}^{(n+1)s}.$$

Similarly, let Z denote the set of inputs (times/locations and derivative orders) used by the observation operator $\mathcal{H}$. Because $(\mathbf{y}_*, \mathcal{H}[y])$ are jointly Gaussian, the posterior is

$$\mathbf{y}_* \mid \mathbf{z} \;\sim\; \mathrm{N}\big(\mu_*, \Sigma_*\big), \quad \mu_* = K_{*Z}\,(K_{ZZ} + R)^{-1}\,\mathbf{z}, \quad \Sigma_* = K_{**} - K_{*Z}\,(K_{ZZ} + R)^{-1}\,K_{Z*},$$

where the blocks are obtained by differentiating the kernel:

$$\begin{aligned}
\big[K_{**}\big]_{(k,i),(l,j)} &= \frac{\partial^{k+l}}{\partial u^k\, \partial v^l} K(u_i, v_j), \\
\big[K_{*Z}\big]_{(k,i),\,r} &= \mathbb{E}\big[y^{(k)}(u_i)\,(\mathcal{H}[y])_r\big] \quad \text{(differentiate } K \text{ under } \mathcal{H}), \\
K_{ZZ} &= \mathbb{E}\big[\mathcal{H}[y]\,\mathcal{H}[y]^\top\big] \quad \text{(apply } \mathcal{H} \text{ to both kernel arguments).}
\end{aligned}$$

Linear map from derivatives to Q. Define the *block-row* linear operator $L_* \in \mathbb{R}^{s\times(n+1)s}$ acting pointwise as

$$\big(L_*\,\mathbf{y}_*\big)_i \;=\; \sum_{k=0}^{n} a_k(u_i)\,y^{(k)}(u_i), \qquad i=1,\dots,s.$$

Then the posterior for $\mathbf{Q}_* := Q(U_*)$ is Gaussian:

$$\boxed{\mathbf{Q}_* \mid \mathbf{z} \;\sim\; \mathrm{N}\Big(\underbrace{L_*\,\mu_*}_{\text{Bayes estimator}}\;,\;\underbrace{L_*\,\Sigma_*\,L_*^\top}_{\text{posterior covariance}}\Big).}$$

Equivalently, the *pointwise* Bayes estimator is

$$\hat{Q}(u_i) \;=\; \sum_{k=0}^{n} a_k(u_i)\,\mu_*^{(k)}(u_i).$$

3.4 Matrix Approach

For many users it is convenient to think entirely in matrices at a grid U_*. Let $D_k \in \mathbb{R}^{s\times s}$ denote the linear map that returns the k^{th} derivative of the GP *at* U_* when applied to the latent vector $y(U_*)$.[1] Let $A_k = \operatorname{diag}\big(a_k(U_*)\big)$. Define the assembled operator

$$L \;=\; \sum_{k=0}^{n} A_k\,D_k \;\in\; \mathbb{R}^{s\times s_{\text{stack}}},$$

where D_k selects the k-th derivative block from $\mathbf{y}_*$ (or, equivalently, we realize L_* directly as above). Then

$$\hat{\mathbf{Q}} \;=\; L\,\mu_*, \qquad \operatorname{Cov}(\mathbf{Q}) \;=\; L\,\Sigma_*\,L^\top.$$

This is the implementation once we have μ_*, Σ_*.

3.5 Observations (including boundary conditions included)

Common choices of $\mathcal{H}$ and corresponding kernel manipulations:

- **Noisy point values:** $z_i = y(t_i) + \epsilon_i$. Then $K_{ZZ} = K(t_i, t_j)$, $R = \operatorname{diag}(\sigma_i^2)$, and K_{*Z} uses $\partial_u^k K(u_i, t_j)$.
- **Noisy derivatives:** $z_i = y^{(r_i)}(t_i) + \epsilon_i$. Differentiate kernel in the t argument accordingly.
- **Exact boundary conditions:** $y(b_\ell) = c_\ell$ or $y^{(r)}(b_\ell) = c$. Encode as rows of $\mathcal{H}$ with a tiny noise variance (e.g. 10^{-10}) so that conditioning stays numerically stable.
- **Linear functionals:** $z_i = \int w_i(t) y^{(r_i)}(t)\,\mathrm{d}t + \epsilon_i$. Apply $\mathcal{H}$ to the kernel by pushing the integral/derivative through K.

3.6 Kernels and differentiability

We need K differentiable up to order n in each argument. Examples:

- Squared Exponential (RBF): infinitely differentiable.
- Matérn ν: k-th derivatives exist in mean-square for $k < \nu$. For an n-th order ODE, choose $\nu > n + 1/2$ for comfortable margins.

3.7 Practical Implementation

1. Choose order n and coefficient functions $a_k(u)$.
2. Choose kernel K with sufficient smoothness.
3. Specify linear observations $\mathcal{H}$ and noise R (including boundary conditions).
4. Build kernel blocks by differentiating K as required to form K_{ZZ}, K_{*Z}, K_{**}.
5. Compute μ_*, Σ_* via Gaussian conditioning.
6. Evaluate $\hat{\mathbf{Q}} = L_*\mu_*$ and (optionally) $\operatorname{Cov}(\mathbf{Q}) = L_*\Sigma_*L_*^\top$.

[1] In exact GP algebra, derivatives are obtained via differentiated kernels against observations; D_k here is a mnemonic for "apply the k-th derivative with respect to the first kernel argument and evaluate at U_*". In practice we never build D_k explicitly; we build the corresponding kernel blocks.

3.8 Example: 3'rd Order Linear ODE

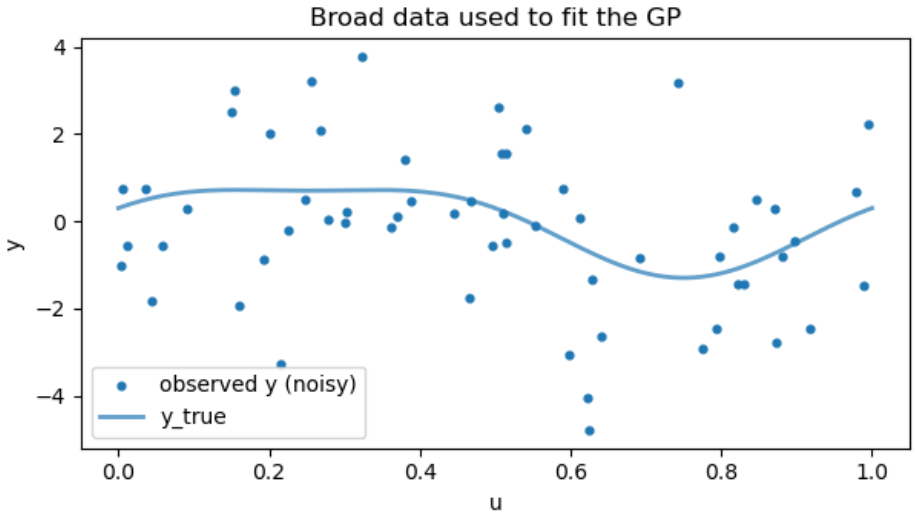

Figure 1: The data and the fit Gaussian Process for the 3'rd order ODE template scenario.

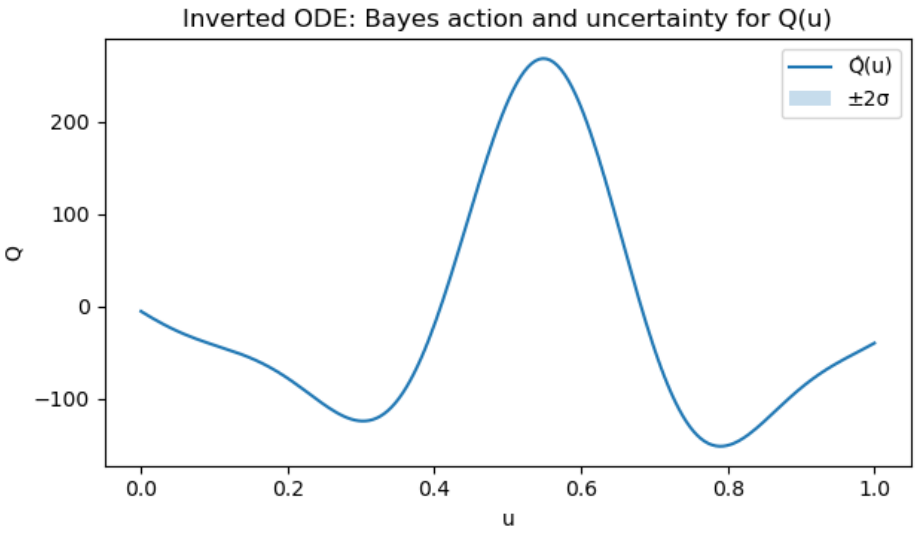

Figure 2: Bayes action and Uncertainty for the 3'rd order linear ODE.

4 Linear k'th Order PDEs

4.1 Setup: multi-index operators on $\Omega \subset \mathbb{R}^d$

Let $\Omega \subset \mathbb{R}^d$ be open with boundary $\partial\Omega$. For a multi-index $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$, denote $D^\alpha \equiv \partial^{|\alpha|}/\partial x_1^{\alpha_1} \circ \partial x_d^{\alpha_d}$, with order $|\alpha| = \sum_j \alpha_j$. Consider a (possibly vector-valued) field $y : \Omega \to \mathbb{R}^p$ and a linear differential operator of order m:

$$(\mathcal{L}y)(x) \;=\; \sum_{|\alpha| \le m} A_\alpha(x)\, D^\alpha y(x), \qquad x \in \Omega,$$

where $A_\alpha(x) \in \mathbb{R}^{q \times p}$ are known coefficient matrices, and $Q : \Omega \to \mathbb{R}^q$ is the unknown source/forcing term:

$$\mathcal{L}y = Q \quad \text{in } \Omega, \qquad \mathcal{B}y = b \quad \text{on } \partial\Omega,$$

with $\mathcal{B}$ a linear boundary operator (e.g. Dirichlet/Neumann/Robin) and known boundary data b.

Prior. Place a (possibly vector-valued) GP prior on y:

$$y \sim \mathrm{GP}\big(0,\, K\big), \quad K : \Omega \times \Omega \to \mathbb{R}^{p\times p},$$

with mean-square derivatives up to order m in each argument. Popular choices include separable kernels $K(x,x') = k(x,x') \otimes \Sigma$ for $\Sigma \in \mathbb{R}^{p\times p}$, or anisotropic products $k(x,x') = \prod_{j=1}^{d} k_j(x_j, x'_j)$.

Observations (interior + boundary, general linear). Let $\mathcal{H}$ be any linear operator mapping y to $\mathbb{R}^m$: pointwise values $y(x_i)$, derivatives $D^\beta y(x_i)$, integrals $\int_\Omega w(x)^\top y(x)\,dx$, or boundary traces $(\mathcal{T}y)(\xi_i)$ at $\xi_i \in \partial\Omega$ (Dirichlet/Neumann/Robin). Assume

$$\mathbf{z} \;=\; \mathcal{H}[y] + \varepsilon, \qquad \varepsilon \sim \mathrm{N}(0, R).$$

Noiseless constraints (e.g. exact boundary data) are obtained by taking very small diagonal entries in R for the corresponding rows.

4.2 Decision problem and Bayes action for $Q = \mathcal{L}y$

Under squared loss over Ω,

$$\mathcal{R}(\hat{Q}) = \mathbb{E}\left[\int_\Omega \|\mathcal{L}y(x) - \hat{Q}(x)\|_2^2\,dx \,\middle|\, \mathbf{z}\right],$$

the pointwise Bayes action is

$$\boxed{\hat{Q}(x) \;=\; \mathbb{E}\big[(\mathcal{L}y)(x) \mid \mathbf{z}\big] \;=\; \sum_{|\alpha|\le m} A_\alpha(x)\,\mathbb{E}\big[D^\alpha y(x) \mid \mathbf{z}\big].}$$

Uncertainty is given by the posterior covariance

$$\mathrm{Cov}\big(Q(x), Q(x') \mid \mathbf{z}\big) \;=\; \mathrm{Cov}\big((\mathcal{L}y)(x), (\mathcal{L}y)(x') \mid \mathbf{z}\big).$$

4.3 Block-Gaussian conditioning with multi-indices

Let $X_* = \{x_1, \dots, x_s\} \subset \Omega$ be query points for Q. Stack the needed derivatives of y at X_* up to order m:

$$\mathbf{y}_* \;=\; \big[\, D^\alpha y(X_*)\,\big]_{|\alpha|\le m} \;\in\; \mathbb{R}^{p(N_\alpha)s}, \quad N_\alpha = \#\{\alpha : |\alpha| \le m\}.$$

Let Z denote the set of functionals used by $\mathcal{H}$. By linearity, $(\mathbf{y}_*, \mathcal{H}[y])$ are jointly Gaussian with

$$\mathbf{y}_* \mid \mathbf{z} \sim \mathrm{N}(\mu_*, \Sigma_*), \qquad \mu_* = K_{*Z}(K_{ZZ}+R)^{-1}\mathbf{z}, \quad \Sigma_* = K_{**} - K_{*Z}(K_{ZZ}+R)^{-1}K_{Z*}.$$

The blocks are obtained by differentiating the kernel with respect to multi-indices:

$$\begin{aligned}
\big[K_{**}\big]_{(\alpha,i),(\beta,j)} &= D_x^\alpha D_{x'}^\beta K(x_i, x_j),\\
\big[K_{*Z}\big]_{(\alpha,i),\,r} &= \mathbb{E}\big[D^\alpha y(x_i)\,(\mathcal{H}[y])_r\big] \;=\; \mathcal{H}_{x'}\big[D_x^\alpha K(x_i, x')\big],\\
K_{ZZ} &= \mathbb{E}\big[\mathcal{H}[y]\,\mathcal{H}[y]^\top\big] \;=\; \mathcal{H}_x\mathcal{H}_{x'}\big[K(x,x')\big].
\end{aligned}$$

Here $\mathcal{H}_x$ and $\mathcal{H}_{x'}$ denote applying $\mathcal{H}$ to the first/second kernel argument, respectively; boundary traces apply via standard trace operators (e.g. Dirichlet $y|_{\partial\Omega}$, Neumann $\partial_n y|_{\partial\Omega}$).

Linear map to Q. Define L_* so that, for each x_i,

$$(L_*\mathbf{y}_*)_i \;=\; \sum_{|\alpha|\le m} A_\alpha(x_i)\,D^\alpha y(x_i).$$

Then

$$\boxed{\mathbf{Q}_* \mid \mathbf{z} \sim \mathrm{N}\big(\, L_*\mu_*,\; L_*\Sigma_*L_*^\top\big), \qquad \hat{\mathbf{Q}} = L_*\mu_*.}$$

4.4 Boundary operators $\mathcal{B}$ as observations

Dirichlet, Neumann, and Robin (or mixed) conditions are all linear:

$$\mathcal{B}y = B_0 y + \sum_{j=1}^{d} B_j \partial_{x_j} y \quad \text{on } \partial\Omega,$$

with matrix-valued B_j possibly depending on the boundary location. Encode these as rows of $\mathcal{H}$ at boundary points (or boundary integrals) with tiny noise for exactness. Their covariance contributions come from tracing the kernel to $\partial\Omega$, e.g. $\mathbb{E}[y(\xi)\,y(\xi')^\top] = K(\xi,\xi')$ for $\xi,\xi' \in \partial\Omega$, $\mathbb{E}[\partial_n y(\xi)\,y(\xi')^\top] = \partial_{n_\xi} K(\xi,\xi')$, etc.

4.5 Computational structure and separable kernels

If K is separable across dimensions, $k(x, x') = \prod_{j=1}^{d} k_j(x_j, x'_j)$, then many derivative blocks factorize, and on tensor grids we can exploit Kronecker algebra for K_{**}, K_{*Z}, K_{ZZ}. For vector-valued y with $K(x, x') = k(x, x') \otimes \Sigma$, use block-diagonalization via the eigen-decomposition of Σ.

4.6 Two templates

(i) Poisson For scalar y and

$$\Delta y(x) = Q(x) \quad \text{in } \Omega,$$

we have $m = 2$ with $A_{e_j+e_j}(x) = -\kappa(x)$ (second derivatives), cross-terms from spatially varying κ via product rule if desired Build L_* using second-order derivatives $D^\alpha y$ for $|\alpha| \leq 2$. Dirichlet or Neumann boundaries become $\mathcal{H}$-rows on $\partial\Omega$.

(ii) Advection–Diffusion–Reaction.

$$-\nabla \circ (D(x)\nabla y) + \mathbf{v}(x) \circ \nabla y + \rho(x)\, y = Q.$$

Here $m = 2$; include first- and second-order derivative blocks for ∇y and $\nabla\nabla y$, assemble L_* with the appropriate $A_\alpha(x)$ derived from $D, \mathbf{v}, \rho$.

4.7 Practical Implementation (PDE)

1. Choose m, coefficient matrices $A_\alpha(x)$, and boundary operator $\mathcal{B}$.
2. Pick K with mean-square derivatives up to m (e.g. RBF; Matérn with $\nu > m + d/2$).
3. Specify $\mathcal{H}$ including boundary rows; set R (tiny variance for exact constraints).
4. Build kernel blocks by applying D_x^α, $D_{x'}^\beta$, traces, and integrals to K.
5. Compute μ_*, Σ_* and return $\hat{\mathbf{Q}} = L_*\mu_*$, $\mathrm{Cov}(\mathbf{Q}_*) = L_*\Sigma_* L_*^\top$.

4.8 Example: Poisson Boundary Conditions

We illustrate the example on the Poisson equation. We sample 250 values uniformly from $[0, 1] \times [0, 1]$ evaluate the following formula on the values:

$$f(x, y) = 5 * \sin(2\pi x)\cos(2\pi y) + 2 * \sin(6\pi xy) \tag{6}$$

and add iid Gaussian noise $z = f(x, y) + \mathrm{N}(0, 2)$. An image depicting the data is available in Figure 3.

We then fit a Gaussian process with radial basis function kernel (RBF) given by:

$$k(a, b; \sigma^2, s) = \sigma^2 \exp[-\frac{||a - b||^2}{2s^2}] \tag{7}$$

This is fit to the data using a zero mean GP with RBF kernel and we get a posterior process $y|z \sim \mathrm{N}(\mu_*, \sigma_*)$.

Laplacian coefficients are built over a 50×50 uniform grid on $[0, 1] \times [0, 1]$ using finite differences. This giving us an approximation to the Laplacion $\hat{\Delta}$; which is then applied to the original GP to give $\hat{\Delta}(y|z) \sim \mathrm{N}(\hat{\Delta}(\mu), \hat{\Delta}\sigma_*\hat{\Delta}^\top)$. The mean of this is the Bayes risk under squared error loss for the source term. A plot illustrating the mean is given in Figure 4. We note that by Theorem 2 this is equivalent to the 0-1 loss minimizer, and by Theorem 1 it is an approximation on the grid projection of the Hilbert space Bayes action for the Laplacian operator (as was equivalent in Section /3.8).

For uncertainty estimates we take the trace of the discretized operator $\Lambda_*\sigma_*\Lambda_*^\top$ and evaluate the $\mu_* \pm 2\hat{\sigma}$. This is provide on the discretized $[0, 1] \times [0, 1]$ in Figure 5.

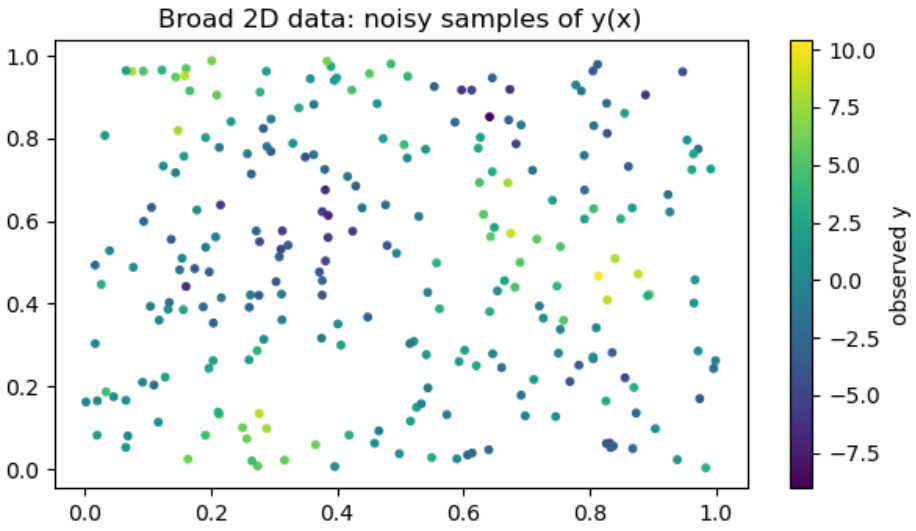

Figure 3: The data distributed on a 2 dimensional grid which the Gaussian process $y|z$ on $\mathbb{R}^2$ that the Gaussian Process is fit to.

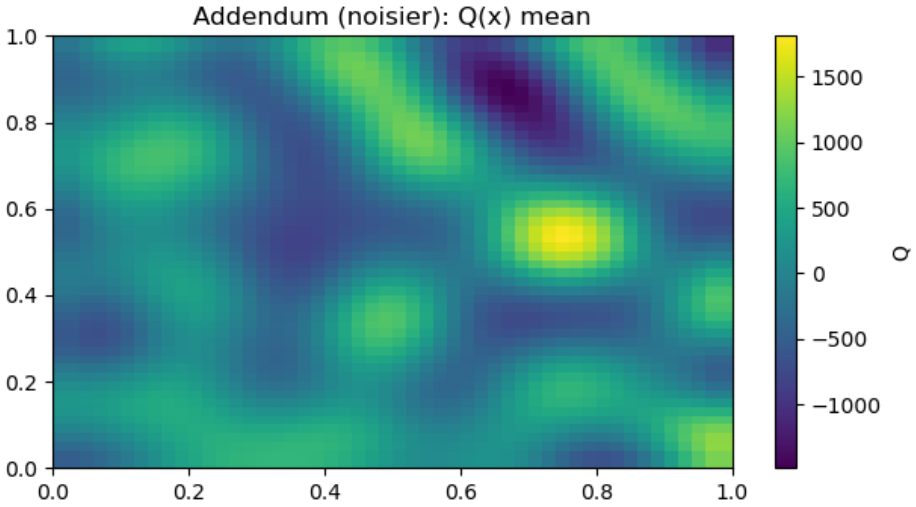

Figure 4: Bayes action of of the resultant Dirichlet loss resulting from taking the mean of the Gaussian process after applying the Laplacian operator.

5 Performance guarantees for nonlinear operators via Frechet derivatives

Let $\mathcal{Y}, \mathcal{Q}$ be Hilbert spaces equipped with inner products $\langle \cdot, \cdot \rangle_{\mathcal{Y}}$ and $\langle \cdot, \cdot \rangle_{\mathcal{Q}}$, respectively. Suppose we are interested in a (generally nonlinear) operator $\mathcal{N} : \mathcal{Y} \to \mathcal{Q}$ and the quantity $Q = \mathcal{N}(y)$. From the Gaussian process (GP) posterior over the latent field y given data z we have

$$y \mid z \sim \mathrm{N}(\mu_\star, \Sigma_\star), \qquad \widehat{Q} := \mathrm{N}(\mu_\star).$$

Writing the posterior error as $e := y - \mu_\star$, we expand $\mathcal{N}$ around $\mu_\star$ using Frechet calculus.

5.1 First-order linearization and Bayes risk

If $\mathcal{N}$ is Frechet differentiable at $\mu_\star$ with derivative $D\mathcal{N}[\mu_\star] \in \mathcal{L}(\mathcal{Y}, \mathcal{Q})$, then

$$\mathcal{N}(\mu_\star + e) = \mathcal{N}(\mu_\star) + D\mathcal{N}[\mu_\star](e) + R_2(e), \qquad \|R_2(e)\|_{\mathcal{Q}} = o(\|e\|_{\mathcal{Y}}). \tag{8}$$

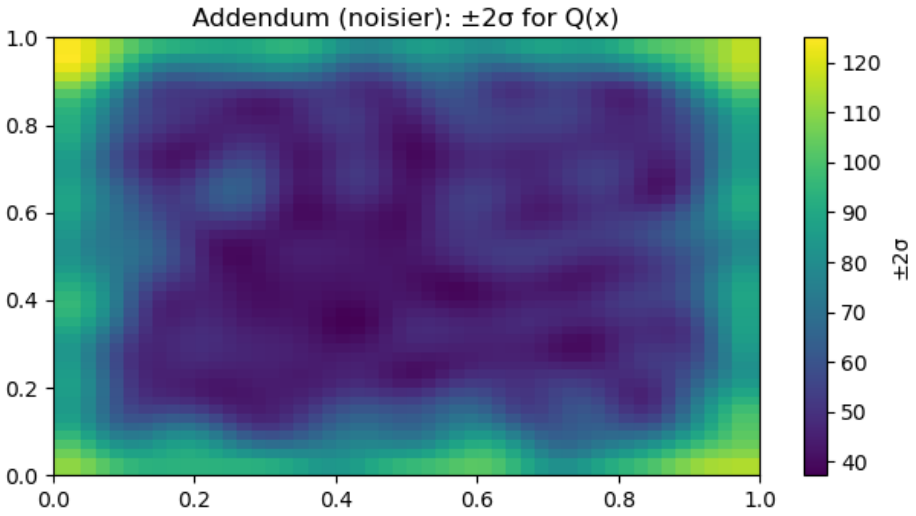

Figure 5: Uncertainty around the Bayes action after application of the Dirichlet operator.

Under squared loss the first-order approximation of the posterior Bayes risk is

$$\mathbb{E}\big[\|Q-\widehat{Q}\|_{\mathcal{Q}}^2 \,\big|\, z\big] \;\approx\; \mathbb{E}\big[\|D\mathcal{N}[\mu_\star](e)\|_{\mathcal{Q}}^2 \,\big|\, z\big] \;=\; \mathrm{tr}\Big(D\mathcal{N}[\mu_\star]\,\Sigma_\star\,D\mathcal{N}[\mu_\star]^*\Big). \tag{9}$$

Hence a conservative spectral bound is

$$\mathbb{E}\big[\|Q-\widehat{Q}\|_{\mathcal{Q}}^2 \,\big|\, z\big] \;\le\; \|D\mathcal{N}[\mu_\star]\|_{\mathrm{op}}^2\,\mathrm{tr}(\Sigma_\star), \tag{10}$$

where $\|\cdot\|_{\mathrm{op}}$ denotes the operator norm and $\mathrm{tr}(\cdot)$ the Hilbert–Schmidt trace.

5.2 High-probability certificate

Let d be the dimension of a finite discretization and define the $(1-\alpha)$-credible ellipsoid $\mathcal{E}_\alpha = \{e:\; e^\top\Sigma_\star^{-1}e \le \chi^2_{d,\alpha}\}$. Then, with probability at least $1-\alpha$,

$$\|Q-\widehat{Q}\|_{\mathcal{Q}}v \;\lesssim\; \|D\mathcal{N}[\mu_\star]\|_{\mathrm{op}}\,\sqrt{\lambda_{\max}(\Sigma_\star)}\,\sqrt{\chi^2_{d,\alpha}} \;+\; \sup_{e\in\mathcal{E}_\alpha}\|R_2(e)\|_{\mathcal{Q}}. \tag{11}$$

5.3 Second-order curvature control

If $\mathcal{N}$ is twice Frechet differentiable in a neighborhood of $\mu_\star$ and $\|D^2\mathcal{N}[y]\|_{\mathrm{op}} \le M$ there, the remainder admits the quadratic bound

$$\|R_2(e)\|_{\mathcal{Q}} \;\le\; \tfrac{1}{2}M\|e\|_{\mathcal{Y}}^2. \tag{12}$$

Taking expectations yields

$$\mathbb{E}\big[\|Q-\widehat{Q}\|_{\mathcal{Q}}^2 \,\big|\, z\big] \;\le\; \|D\mathcal{N}[\mu_\star]\|_{\mathrm{op}}^2\,\mathrm{tr}(\Sigma_\star) \;+\; \tfrac{1}{2}M^2\,\mathbb{E}\|e\|_{\mathcal{Y}}^4 \;+\; \text{cross terms}, \tag{13}$$

where for Gaussian e the fourth moment is determined by $\mathrm{tr}(\Sigma_\star)$ and $\|\Sigma_\star\|_F^2$.

5.4 Lipschitz (global) guarantees

If $\mathcal{N}$ is (locally) Lipschitz with constant L near $\mu_\star$,

$$\|Q-\widehat{Q}\|_{\mathcal{Q}} \le L\|e\|_{\mathcal{Y}}, \qquad \mathbb{E}\big[\|Q-\widehat{Q}\|_{\mathcal{Q}}^2 \,\big|\, z\big] \le L^2\,\mathrm{tr}(\Sigma_\star). \tag{14}$$

Typically one can bound L via the derivative as $L \ge \sup_{y\in\mathcal{U}}\|D\mathcal{N}[y]\|_{\mathrm{op}}$ on a neighborhood $\mathcal{U}$.

5.5 Operator templates

Pointwise nonlinearity $\mathcal{N}(y) = f\circ y$. For smooth $f:\mathbb{R}\to\mathbb{R}$, $[D\mathcal{N}[\mu_\star](h)](x) = f'(\mu_\star(x))\,h(x)$. Hence $\|D\mathcal{N}[\mu_\star]\|_{\mathrm{op}} = \|f'(\mu_\star)\|_{L^\infty}$. If f'' is bounded by M_f, then $M \le \|f''(\mu_\star)\|_{L^\infty}$ in (12).

Bilinear convective term $\mathcal{N}(y) = B(y, y)$. For a bounded bilinear form B, $D\mathcal{N}[y](h) = B(h, y) + B(y, h)$, so $\|D\mathcal{N}[\mu_\star]\|_{\mathrm{op}} \le 2\|B\|_{\mathrm{op}}\|\mu_\star\|$ and $R_2(e) = B(e, e)$ obeys $\|R_2(e)\| \le \|B\|_{\mathrm{op}}\|e\|^2$.

Coefficient-nonlinear elliptic term $\mathcal{N}(y) = -\nabla\cdot\big(a(y)\nabla y\big)$. The derivative is

$$D\mathcal{N}[\mu_\star](h) = -\nabla\cdot\big(a'(\mu_\star)\, h\, \nabla\mu_\star + a(\mu_\star)\, \nabla h\big),$$

leading to bounds (up to domain-dependent constants) of the form $\|D\mathcal{N}[\mu_\star]\|_{\mathrm{op}} \lesssim \|a'(\mu_\star)\|_{L^\infty}\|\nabla\mu_\star\|_{L^\infty} + \|a(\mu_\star)\|_{L^\infty}$.

5.6 Reporting and design

For transparency and reproducibility, we recommend reporting:

1. the first-order covariance $\Sigma_Q^{(1)} := D\mathcal{N}[\mu_\star]\Sigma_\star D\mathcal{N}[\mu_\star]^*$ and its trace/spectral radius;
2. an estimate of $\|D\mathcal{N}[\mu_\star]\|_{\mathrm{op}}$ (e.g., power iteration);
3. a curvature constant M (or an empirical ratio $\|R_2(e)\|/\|D\mathcal{N}[\mu_\star](e)\|$ on samples $e \sim \mathcal{N}(0, \Sigma_\star)$);
4. a high-probability radius r_α from (11).

These quantities also drive active data acquisition by targeting reductions in $\mathrm{tr}(\Sigma_Q^{(1)})$ or the spectral radius of $\Sigma_Q^{(1)}$.

6 Acknowledgements

Code available at the github repo **INSERT LINK HERE** and was generated with the help of GPT5.

7 Notes for me:

Regularity. Interchange of derivatives/expectation and trace operations is justified under standard GP regularity; ensure K is sufficiently smooth and the trace/normal derivatives are defined (e.g. via Sobolev embeddings).

Learning. Hyperparameters of K and any unknown coefficients in A_α can be learned by marginal likelihood or hierarchical Bayes using K_{ZZ}.

Vector fields. For $p, q > 1$, use matrix-valued kernels or LMC constructions; the formulas above apply verbatim with matrix blocks.

Nonconstant coefficients. All formulas above already allow $a_k = a_k(u)$ varying with u (they only enter via the diagonal A_k at test points).

Model selection. We can learn kernel hyperparameters by maximizing the marginal likelihood of $\mathbf{z}$ (standard GP training); $\hat{\mathbf{Q}}$ then updates accordingly. Possibly extend to quantification of overfitting of data in this context and associated Bayes Risk perturbations.

Design/active sampling. Since we have $\mathrm{Cov}(\mathbf{Q})$, we can design new measurements to reduce uncertainty in Q, e.g. choose new times to minimize $\mathrm{tr}(L_*\Sigma_* L_*^\top)$.

A Proof of Theorem 1

Statement 1. *Let $\mathcal{H}$ be a real, separable Hilbert space. Let $\Psi : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ be a loss function, and let Q be a random element in $\mathcal{H}$ drawn from a posterior measure given data z. Let $\{H_n\}_{n=1}^\infty$ be an increasing sequence of finite-dimensional subspaces such that $\overline{\bigcup H_n} = \mathcal{H}$, and let $P_n : \mathcal{H} \to H_n$ be the corresponding orthogonal projection operators.*

Define the global expected posterior risk $R(a) = \mathbb{E}_{Q|z}[\Psi(a, Q)]$ and the finite-dimensional projected risk $R_n(a) = \mathbb{E}_{Q|z}[\Psi(a, P_n Q)]$. Assume the following conditions hold:

1. ***Growth Bound:*** *There exist $C > 0$ and $p \ge 1$ such that $|\Psi(a, h)| \le C(1 + \|a\|^p + \|h\|^p)$, and $\mathbb{E}_{Q|z}[\|Q\|^p] < \infty$.*

2. ***Lipschitz Continuity in the State:*** *There exists $L > 0$ such that for all $a, h, h' \in \mathcal{H}$, $|\Psi(a, h) - \Psi(a, h')| \leq L\|h - h'\|$.*
3. ***Continuity in the Action:*** *For any fixed $h \in \mathcal{H}$, the map $a \mapsto \Psi(a, h)$ is continuous.*
4. ***Coercivity:*** *$R(a) \to \infty$ as $\|a\| \to \infty$.*
5. ***Uniform Convexity:*** *$R(a)$ is uniformly convex on $\mathcal{H}$.*

Then, for each n, there exists a unique minimizer $a_n = \arg\min_{a \in H_n} R_n(a)$. Furthermore, the sequence (a_n) converges strongly to the unique global Bayes action $a^* = \arg\min_{a \in \mathcal{H}} R(a)$; that is, $\lim_{n\to\infty} \|a_n - a^*\| = 0$.

Proof. **Part 1: Existence and Uniqueness in H_n**
Because H_n is a finite-dimensional subspace, it is isomorphic to $\mathbb{R}^d$ for some d. By the Dominated Convergence Theorem (DCT)—justified by the growth bound (Condition 1) and the continuity of Ψ in its first argument (Condition 3)—the functional $R_n(a)$ is continuous on H_n.

Because $R_n(a)$ inherits coercivity (Condition 4) from $R(a)$, the sub-level sets $\{a \in H_n : R_n(a) \leq C\}$ are closed and bounded. In finite dimensions, closed and bounded sets are compact. By the Weierstrass Extreme Value Theorem, the continuous functional $R_n(a)$ achieves its minimum on H_n. Furthermore, because uniform convexity (Condition 5) implies strict convexity, this minimizer is unique. Thus, we may uniquely define:

$$a_n = \arg\min_{a \in H_n} R_n(a) \tag{15}$$

Part 2: Boundedness of the Sequence of Minimizers
Evaluate the risk at the origin, $R_n(0) = \mathbb{E}_{Q|z}[\Psi(0, P_n Q)]$. Because $P_n Q \to Q$ strongly for almost every realization of Q, and because the loss is bounded by the integrable polynomial envelope (Condition 1), the DCT implies that $\lim_{n\to\infty} R_n(0) = R(0)$.

Because converging sequences are bounded, there exists a constant $M > 0$ such that $R_n(0) \leq M$ for all n. By the definition of the minimizer a_n, we have:

$$R_n(a_n) \leq R_n(0) \leq M \tag{16}$$

Because R_n is uniformly coercive across all n, the bound on the functional's minimum implies a bound on the norm of the minimizers. Therefore, there exists some $K > 0$ such that $\|a_n\| \leq K$ for all n. The sequence (a_n) is therefore bounded in $\mathcal{H}$.

Part 3: Weak Convergence to the Global Minimizer
Because bounded sequences in Hilbert spaces are weakly precompact (Banach-Alaoglu Theorem), the sequence (a_n) has a weakly convergent subsequence (a_{n_k}) that converges to some limit $a^* \in \mathcal{H}$:

$$a_{n_k} \rightharpoonup a^* \tag{17}$$

We must show that a^* is the global minimizer of $R(a)$. Let $h \in \mathcal{H}$ be an arbitrary test element. Define $h_k = P_{n_k} h$. By the properties of orthogonal projections, $h_k \to h$ strongly in $\mathcal{H}$. Since a_{n_k} is the minimizer in H_{n_k}, it follows that:

$$R_{n_k}(a_{n_k}) \leq R_{n_k}(h_k) \tag{18}$$

We now take the limit infimum ($\liminf$) as $k \to \infty$. Using the Lipschitz bound (Condition 2), we know $|R_{n_k}(a) - R(a)| \leq L\mathbb{E}_{Q|z}[\|P_{n_k} Q - Q\|]$. By the DCT, this difference vanishes uniformly on bounded sets. Thus, we can replace R_{n_k} with R in the limit.

Because $R(a)$ is continuous and convex, it is Weakly Lower Semicontinuous (WLSC). Therefore:

$$R(a^*) \leq \liminf_{k\to\infty} R(a_{n_k}) = \liminf_{k\to\infty} R_{n_k}(a_{n_k}) \tag{19}$$

Chaining the inequalities together:

$$R(a^*) \leq \liminf_{k\to\infty} R_{n_k}(a_{n_k}) \leq \lim_{k\to\infty} R_{n_k}(h_k) = R(h) \tag{20}$$

Since $R(a^*) \leq R(h)$ for all $h \in \mathcal{H}$, a^* is a global minimizer. By the strict convexity of $R(a)$, the global minimizer is unique. Because every weakly convergent subsequence of the bounded sequence (a_n) must converge to the same unique limit a^*, the entire sequence converges weakly: $a_n \rightharpoonup a^*$.

Part 4: Strong Convergence
Weak convergence ($a_n \rightharpoonup a^*$) does not automatically imply that the norm distance $\|a_n - a^*\|$ shrinks to zero. However, we assumed $R(a)$ is uniformly convex (Condition 5).

For a uniformly convex functional, if a sequence weakly converges to the minimizer ($a_n \rightharpoonup a^*$) and the functional values converge ($R(a_n) \to R(a^*)$), then the sequence converges strongly.

By taking $h = a^*$ in the inequality from Part 3, we have $h_n = P_n a^* \to a^*$ strongly, and:

$$R(a^*) \leq \liminf_{n\to\infty} R(a_n) \leq \limsup_{n\to\infty} R(a_n) \leq \lim_{n\to\infty} R_n(P_n a^*) = R(a^*) \tag{21}$$

This squeezes the sequence, proving that $\lim_{n\to\infty} R(a_n) = R(a^*)$. Due to the uniform convexity of the risk functional R, the weak convergence and the convergence of the functional values jointly force strong convergence in the Hilbert space norm:

$$\lim_{n\to\infty} \|a_n - a^*\| = 0 \tag{22}$$

Which concludes the proof. □

B Proof of Theorem 2

Statement 2 (0-1 vs. squared-error Bayes equivalence in Gaussian Process Measures on Hilbert spaces)**.** *Let $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ be a real, separable Hilbert space equipped with the norm $\|\cdot\|$. Let Q be a random element in $\mathcal{H}$ whose posterior distribution given data z is Gaussian:*

$$Q \mid z \sim \mathcal{N}(\mu, \mathcal{C}) \tag{23}$$

where $\mu \in \mathcal{H}$ is the posterior mean, and $\mathcal{C} : \mathcal{H} \to \mathcal{H}$ is a self-adjoint, positive, trace-class covariance operator.

Let $\{H_n\}_{n=1}^{\infty}$ be an increasing sequence of finite-dimensional closed subspaces such that $H_n \uparrow \mathcal{H}$, and let $P_n : \mathcal{H} \to H_n$ denote the orthogonal projection operator onto H_n. For a localized action $a_n \in H_n$ and a fixed radius $\tau > 0$, define the sequence of finite-dimensional projected risk minimization problems under the L_2 loss and the $0-1$ small-ball loss respectively:

$$a^*_{2,n} = \arg\min_{a_n \in H_n} \mathbb{E}_{Q|z}\left[\|P_n Q - a_n\|^2 \;\middle|\; z\right] \tag{24}$$

$$a^*_{01,n} = \arg\min_{a_n \in H_n} \mathbb{E}_{Q|z}\left[\mathbb{1}\{\|P_n Q - a_n\| > \tau\} \mid z\right] \tag{25}$$

Then, for every $n \geq 1$, the following properties hold:

1. ***Subspace Equivalence:*** *The unique minimizer on every finite-dimensional subspace under both losses is exactly the projected posterior mean:*
$$a^*_{2,n} = a^*_{01,n} = P_n \mu \tag{26}$$
2. ***Infinite-Dimensional Limit:*** *As $n \to \infty$, the sequence of subspace optimal actions converges strongly in $\mathcal{H}$ to the unique global infinite-dimensional Bayes action $a^* = \mu$:*
$$\lim_{n\to\infty} \|a^*_{2,n} - \mu\| = 0 \quad \textit{and} \quad \lim_{n\to\infty} \|a^*_{01,n} - \mu\| = 0 \tag{27}$$

Proof. **Part 1: Solving the Projected Subspace Problems**
Because $P_n : \mathcal{H} \to H_n$ is a bounded linear operator, the projected random variable $P_n Q \mid z$ is a tightly supported, finite-dimensional Gaussian element in the subspace H_n. Its induced Radon measure within H_n is completely characterized by:

$$P_n Q \mid z \sim \mathcal{N}(P_n \mu, P_n \mathcal{C} P_n) \tag{28}$$

First, we resolve the optimal finite-dimensional action $a^*_{2,n}$ for the L_2 risk. Expanding the expected error using the linearity of expectation and the orthogonality of the mean centering yields:

$$\mathbb{E}_{Q|z}\left[\|P_n Q - a_n\|^2 \;\middle|\; z\right] = \mathbb{E}_{Q|z}\left[\|P_n Q - P_n \mu\|^2 \;\middle|\; z\right] + \|P_n \mu - a_n\|^2 \tag{29}$$

The first expectation on the right-hand side evaluates identically to the subspace trace $\mathrm{tr}(P_n \mathcal{C} P_n)$, which is a constant independent of a_n. The remaining functional $\|P_n \mu - a_n\|^2$ is strictly convex and possesses a unique global minimum if and only if its value is zero. Hence, we uniquely obtain:

$$a^*_{2,n} = P_n \mu \tag{30}$$

Next, we resolve the optimal finite-dimensional action $a^*_{01,n}$ under the 0-1 loss. Minimizing the expected posterior 0-1 risk is mathematically equivalent to maximizing the probability that the approximation error falls within a closed ball of radius τ centered at the origin of H_n:

$$\arg\min_{a_n \in H_n} \mathbb{E}_{Q|z}\left[\mathbb{1}\{\|P_nQ - a_n\| > \tau\} \mid z\right] = \arg\max_{a_n \in H_n} \mathbb{P}\left(\|P_nQ - a_n\| \leq \tau \mid z\right) \tag{31}$$

$$= \arg\max_{a_n \in H_n} \mathbb{P}\left(P_nQ - a_n \in B_n(0,\tau) \mid z\right) \tag{32}$$

Let $X_n = P_nQ - P_n\mu$. By construction, X_n is a centered (mean $= 0$) Gaussian random variable restricted to the finite-dimensional space H_n. We rewrite the probability functional as a shift expression:

$$\mathbb{P}\left(X_n + P_n\mu - a_n \in B_n(0,\tau) \mid z\right) \tag{33}$$

The closed ball $B_n(0,\tau) \subset H_n$ is a symmetric and convex set. By **Anderson's Shift Inequality**, for any centered, symmetric, unimodal Gaussian measure on a finite-dimensional vector space, translating a symmetric convex target away from the origin strictly reduces its measure:

$$\mathbb{P}(X_n + v \in B_n(0,\tau)) < \mathbb{P}(X_n \in B_n(0,\tau)) \quad \forall v \in H_n \setminus \{0\} \tag{34}$$

This probability achieves its strictly unique maximum when the translation displacement vector vanishes entirely, meaning $v = P_n\mu - a_n = 0$. Consequently, the unique optimal action on the subspace is given by:

$$a^*_{01,n} = P_n\mu \tag{35}$$

Thus, for any finite dimension n, both loss criteria dictate identical decisions on the subspace: $a^*_{2,n} = a^*_{01,n} = P_n\mu$.

Part 2: Infinite-Dimensional Convergence
We now embedded these finite-dimensional choices within the global asymptotic framework. Because the nested sequence of finite-dimensional closed subspaces converges upward to the full space ($\overline{\bigcup H_n} = \mathcal{H}$), the sequence of orthogonal projection operators P_n converges strongly to the identity operator I on $\mathcal{H}$. That is, for any deterministic element $v \in \mathcal{H}$:

$$\lim_{n\to\infty} \|P_nv - v\| = 0 \tag{36}$$

Since the true global infinite-dimensional posterior mean satisfies $\mu \in \mathcal{H}$, substituting μ into the strong convergence property of the projection sequence immediately yields:

$$\lim_{n\to\infty} a^*_{2,n} = \lim_{n\to\infty} P_n\mu = \mu \tag{37}$$

$$\lim_{n\to\infty} a^*_{01,n} = \lim_{n\to\infty} P_n\mu = \mu \tag{38}$$

Because the exact global risk functionals over the full space $\mathcal{H}$ are uniquely minimized by the infinite-dimensional Bayes action $a^* = \mu$, the sequence of finite-dimensional actions under both loss functions converges strongly in the norm topology of the Hilbert space to the global Bayes action:

$$\lim_{n\to\infty} \|a^*_{2,n} - a^*\| = 0 \quad \text{and} \quad \lim_{n\to\infty} \|a^*_{01,n} - a^*\| = 0 \tag{39}$$

This completes the proof. □

C Proof of Theorem 3

Statement 3 (Decision boundary for two Gaussian-process posteriors)**.** *Let $\mathcal{H}$ be a real, separable Hilbert space. Let $Y \sim \mathcal{N}(\mu, \mathcal{C})$ and $Y' \sim \mathcal{N}(\mu', \mathcal{C}')$ be Gaussian random elements in $\mathcal{H}$ with means $\mu, \mu' \in \mathcal{H}$ and covariance operators $\mathcal{C}, \mathcal{C}'$ (self-adjoint, positive, trace-class on the closed span of their supports). For any $a \in \mathcal{H}$, define the small-ball* 0–1 *risks*

$$R^Y_{01,\tau}(a) := \mathbb{E}[\mathbf{1}\{\|Y - a\| > \tau\}], \qquad R^{Y'}_{01,\tau}(a) := \mathbb{E}[\mathbf{1}\{\|Y' - a\| > \tau\}].$$

Define the decision boundary *as the zero-level set of the limiting risk difference,*

$$\mathcal{B} := \Big\{a \in \mathcal{H} : \lim_{\tau\downarrow 0} \big(R^Y_{01,\tau}(a) - R^{Y'}_{01,\tau}(a)\big) = 0\Big\}.$$

Then $\mathcal{B}$ is a Hilbert manifold in $\mathcal{H}$, and coincides with the zero-level set of the squared-loss risk difference,

$$\mathcal{B} = \Big\{a \in \mathcal{H} : \mathbb{E}\|Y - a\|^2 = \mathbb{E}\|Y' - a\|^2\Big\},$$

and therefore admits the explicit affine description

$$\boxed{\|a-\mu\|^2 - \|a-\mu'\|^2 = \mathrm{tr}(\mathcal{C}') - \mathrm{tr}(\mathcal{C}).}$$

Which is an affine Hilbert manifold in $\mathcal{H}$ locally diffeomorphic to $\mathcal{H}$. In particular, when $\mathcal{C} = \mathcal{C}'$, the boundary is the (codimension-one) bisector hyperplane Hilbert manifold in $\mathcal{H}$

$$\{\, a \in \mathcal{H} : \langle a, \mu' - \mu\rangle = \tfrac{1}{2}(\|\mu'\|^2 - \|\mu\|^2)\,\},$$

i.e. the locus of points equidistant (in $\|\cdot\|$) from μ and μ'.

Proof. The proof proceeds with the following steps:

1. **Squared-loss risks.** For any Gaussian $Z \sim \mathcal{N}(m, \Sigma)$ in $\mathcal{H}$,
$$\mathbb{E}\|Z-a\|^2 = \|m-a\|^2 + \mathrm{tr}(\Sigma).$$
Applying this to Y and Y' yields
$$\mathbb{E}\|Y-a\|^2 - \mathbb{E}\|Y'-a\|^2 = (\|a-\mu\|^2 - \|a-\mu'\|^2) + (\mathrm{tr}(\mathcal{C}) - \mathrm{tr}(\mathcal{C}')).$$
Hence the squared-loss boundary satisfies
$$\|a-\mu\|^2 - \|a-\mu'\|^2 = \mathrm{tr}(\mathcal{C}') - \mathrm{tr}(\mathcal{C}).$$

2. **Equivalence of small-ball and squared-loss risks.** By Theorem 2, for a Gaussian process posterior $Q \sim \mathcal{N}(m, \Sigma)$, the Bayes action minimizing either squared loss $\|Q-a\|^2$ or small-ball 0–1 loss $\mathbf{1}\{\|Q-a\| > \tau\}$ (for fixed $\tau > 0$) is the posterior mean $a^* = m$. Thus, for both (Y, μ) and (Y', μ'), the ordering of risks under the small-ball loss agrees with that under squared loss as $\tau \downarrow 0$.

3. **Equality of risks.** Define
$$\Delta_2(a) := \mathbb{E}\|Y-a\|^2 - \mathbb{E}\|Y'-a\|^2, \qquad \Delta_{01,\tau}(a) := R^{Y}_{01,\tau}(a) - R^{Y'}_{01,\tau}(a).$$
By Step 1, Δ_2 is affine in a, and by Step 2, $\mathrm{sign}(\Delta_{01,\tau}(a)) = \mathrm{sign}(\Delta_2(a))$ for small τ. Consequently,
$$\lim_{\tau \downarrow 0} \Delta_{01,\tau}(a) = 0 \iff \Delta_2(a) = 0.$$
Hence $\mathcal{B}$ equals the affine set described above.

4. **Manifold structure.** Unless $\mu = \mu'$ and $\mathrm{tr}(\mathcal{C}) = \mathrm{tr}(\mathcal{C}')$, the boundary is given by one nontrivial affine equation in $\mathcal{H}$, so it is an affine hyperplane (locally diffeomorphic to $\mathcal{H}$).

□

D More Simulated Examples

References

James O Berger. *Statistical decision theory and Bayesian analysis*. Springer, 2010.

Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10:273–304, 1995.

Jon Cockayne, Chris J. Oates, T. J. Sullivan, and Mark Girolami. Bayesian probabilistic numerical methods. *SIAM Review*, 61(4):756–789, 2019.

Masoumeh Dashti and Andrew M. Stuart. The bayesian approach to inverse problems, 2015. URL `https://arxiv.org/abs/1302.6989`.

Pierpaolo De Blasi, Stefano Favaro, Antonio Lijoi, Ramses H. Mena, Igor Prunster, and Matteo Ruggiero. Are gibbs-type priors the most natural generalization of the dirichlet process? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):212–229, February 2015. ISSN 2160-9292. doi: 10.1109/tpami.2013.217. URL `http://dx.doi.org/10.1109/TPAMI.2013.217`.

Jari Kaipio and Erkki Somersalo. *Statistical and Computational Inverse Problems*. Springer, 2004.

Jari Kaipio, Ville Kolehmainen, Erkki Somersalo, and Marko Vauhkonen. Statistical inversion and monte carlo sampling methods in electrical impedance tomography. *Inverse Problems*, 23(6):P1–P39, 2007.

J F C Kingman. Completely random measures. *Pacific Journal of Mathematics*, 21(1):59–78, 1967. URL `https://msp.org/pjm/1967/21-1/pjm-v21-n1-p06-p.pdf`.

Michael Lavine. Polya trees and random distributions. *The Annals of Statistics*, 20(3):1222–1235, 1992.

Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear pdes. *Journal of Computational Physics*, 378:686–707, 2019.