Absolute Continuity of the Argmax of a Stochastic Process

Evan Dogariu* Boris Hanin*

April 27, 2024

Abstract

We prove the following criterion: a centered, square-integrable stochastic process indexed by $\mathbb{R}^n$ with (bi-)Lipschitz covariance kernel has an absolutely continuous argmax if and only if for all $k \in \{1, \ldots, n\}$ a k-dimensional pointedness condition holds only on a $\mathcal{H}^{k-1}$-null set of indices. The main machinery involves a reduction to the problem of absolute continuity of the pushforward of the metric projection in an abstract Hilbert space, which is solved using geometry and measure theory. As some consequences, we derive uniqueness of the maximum in a novel, more general setting and also prove a GMT statement that stationary sets enjoy absolute continuity of the metric projection.

Contents

*Princeton University

1 Introduction

This is a draft (notice the lack of introduction and everything). In blue I denote the last condition I have to figure out; everything in blue refers to essentially the same phenomenon. In red I write insane madman mutterings to myself :). The subsections with asterisk * are not needed for the main proof of Theorem 1, but might be independently useful.

Here is the list of things that need to be fixed, in my personal order of importance:

- Come up with a dot-product formulation of k-dimensional open half-space-containment property, and replace all the blue stuff with it (including in Corollary 3).
- Show the two measure theory lemmas that Boris said are "just the Cameron-Martin Theorem".
- Figure out Law(ξ) and learn about the continuity of the Karhunen-Loeve transform.
- Prove Lemma 2 (i.e. figure out what regularity remains as $\alpha_j \to \alpha = 1$).
- *Show Lemma 5 (figure out Brian White's strong convexity). Double check the rest of the stationarity talk. try swimming it up (i.e. Φ is weak solution to minimal surface PDE, what then?)
- Confirm we arent sneakily hiding bad points with the $L^2(T, \gamma)$ stuff in the reduction and Theorem 2.
- *Show that $\Phi(T)$ is μ-a.e. proximinal for the proof of Corollary 1.
- Show r-normal-regularity of $f(U)$ for bi-Lipschitz f. Figure out when r-normal-regularity holds, and adjust the proof of Corollary 1 accordingly.
- *try applying to $E = f(\mathbb{R}^n)$ for f adequately "convex" (Alexandrov theorem for second derivatives and this paper for Monge-Ampere measure)
- *Look at applications of the geometry in approximation theory: geometry of locations of best approximations. Is the class of neural networks a pointed set in the set of functions, and what does that mean for function approximation?
- Check if compactness of T is required to apply Karhunen Loeve transform

2 Stochastic Processes

Let $(\Omega, \mathcal{M}, \mathbb{P})$ be a probability space and $(T, \mathcal{O}, \gamma)$ be a metric measure space.

Definition 1 (Stochastic Process)**.** *We call a collection of random variables $W := (W_t)_{t\in T}$ with $W_t : \Omega \to \mathbb{R}$ a* stochastic process indexed by γ*, with the convention that $t \notin \mathrm{spt}(\gamma) \implies W_t \equiv 0$. If each W_t has mean 0, we say that W is* centered*. W may also be viewed as a map from $\Omega \times T \to \mathbb{R}$ sending $(\omega, t) \mapsto W_t(\omega)$; if $W \in L^2(\Omega \times T, \mathbb{P} \otimes \gamma)$ we say that W is* square-integrable*. W is called a* Gaussian process *if $(W_t)_{t\in S}$ is distributed as a Gaussian random vector for any finite index set $S \subseteq T$.*

Definition 2 (Argmax of Stochastic Process)**.** *Let W be a stochastic process indexed by γ. Define the set-valued random variable $Z_W : \Omega \to 2^T$ via*

$$Z_W(\omega) = \arg\max_{t\in T} W_t(\omega)$$

and the argmax measure *ζ on T as*

$$\zeta_W(A) := \mathbb{P}\{\omega \in \Omega : \ Z(\omega) \cap A \neq \emptyset\} \equiv \mathrm{Law}(Z_W)$$

Remark 1. There is prior work establishing the a.s. uniqueness of the maximum of fairly general Levy and Gaussian processes (see Theorem 4 from [1][1]), and so we can expect that Z is single-valued $\mathbb{P}$-a.s. when these conditions hold. However, we will stick to the set-valued definition given in Definition 2, and eventually our analysis will show that uniqueness of Z is a geometric property and occurs under even more general conditions.

2.1 Preliminaries

As alluded to, the goal of this work is to determine the regularity requirements on W and γ in order for the argmax measure to be absolutely continuous (i.e. $\zeta_W \ll \gamma$). Our contribution is a sufficient and (almost) necessary condition in the case $T \subseteq U$ is a finite-dimensional subset of a Hilbert space. First, three quick definitions in order to state the main result.

Definition 3 (Covariance Function)**.** *For a square-integrable stochastic process W, define the* covariance function *$K_W : T \times T \to \mathbb{R}$ sending $(s,t) \mapsto \mathbb{E}[W_s W_t] = \langle W_s, W_t\rangle_{L^2(\Omega,\mathbb{P})}$. Then, K_W is symmetric and positive and is therefore a kernel function (in fact it is the kernel of the covariance operator, see Theorem 10 in [2]).*

Definition 4 (Directional Derivative)**.** *Let $f : U \to V$ be a map between normed spaces and $t \in U$. If for some fixed $u \in U$ the limit*

$$f'(t; u) := \lim_{\delta\to 0} \frac{f(t + \delta u) - f(t)}{\delta}$$

[1] The assumptions are that the process is indexed by a compact set and that the kernel function is continuous, monotonically increasing in the first variable with the second held fixed, and has $\mathbb{E}\|Z\|^2 < \infty$.

exists w.r.t. $\| \cdot \|_V$*, then we say that* f *has an* directional derivative at t in direction u. *If* $f'(t;u)$ *exists for all* $u \in U$*, then* f *is* Gateaux differentiable at t *(if the limit is uniform in* u*, we say that* f *is* Frechet differentiable at t*). For* $t \in U$*, define the* Gateaux differential of f at t *to be the set of directional derivatives, i.e.*

$$Df(t) := \{f'(t;u) : \ u \in U \text{ and } f'(t;u) \text{ exists}\}$$

We let $\operatorname{dom}(Df)(t) := \{u \in U : \ f'(t;u) \text{ exists}\}$ *denote the set of directions in which* f *is differentiable at* t*. Note that* $Df(t), \operatorname{dom}(Df)(t)$ *are closed cones.*

Definition 5 (Hausdorff Measure)**.** *For some* $n \in \mathbb{N} \cup \{0\}$*, we define the* n*-dimensional* Hausdorff measure $\mathcal{H}^n$ *on a metric space in the usual way. For any set* $A \subseteq U$ *(*U *a vector space) and any* $\alpha \geq 0$*,* $\delta > 0$ *we may define*

$$\mathcal{H}^\alpha_\delta(A) := \inf \left\{ \sum_{i \in \mathbb{N}} \operatorname{diam}(A_i)^\alpha : \ \operatorname{diam}(A_i) \leq \delta \text{ and } A \subseteq \bigcup_{i \in \mathbb{N}} A_i \right\},$$

where the infimum is taken over all admissible countable covers of A*. Since this is monotonically nondecreasing as* $\delta \to 0$ *the limit exists, and so for any* A *and* $\alpha \geq 0$ *the following limit exists:*

$$\mathcal{H}^\alpha(A) := \lim_{\delta \to 0} \mathcal{H}^\alpha_\delta(A)$$

This makes $\mathcal{H}^\alpha$ *into a measure on the ambient metric space, called the* Hausdorff measure*, which is invariant to translations (and rotations, if the space is Hilbert) and agrees with the Lebesgue measure if the ambient space is* $\mathbb{R}^\alpha$*,* $\alpha \in \mathbb{N}$*. We can restrict* $\mathcal{H}^\alpha$ *to a subset via* $(\mathcal{H}^\alpha \llcorner S)(A) := \mathcal{H}^\alpha(S \cap A)$ *for all* A*.*

2.2 Main Results

Assumption 1. *Let* W *be a stochastic process indexed by* γ *with covariance function* $K_W : T \times T \to \mathbb{R}$*. We assume the following:*

1. W *is square-integrable and centered.*
2. $\mathcal{H}^n \llcorner T \ll \gamma$ *for some* $n \in \mathbb{N}$ *(i.e.* γ *is finite-dimensional).*
3. $K_W(t, \cdot) : T \to \mathbb{R}$ *is Lipschitz for* γ*-a.e.* $t \in T$ *with*

$$\int_{t \in T} \operatorname{Lip}(K_W(t, \cdot))^2 d\gamma(t) < \infty$$

Assumption 1 will always be assumed, though we will still mention every time we require it. In addition, some straightforward extra regularity is required to prove the reverse direction, which we state in Assumption 2 below:

Assumption 2 (Extra Assumptions for Converse)**.** *In addition to Assumption 1,*

1. *$\gamma \asymp \mathcal{H}^n \llcorner T$ (i.e. γ is exactly n-dimensional and supported on T).*

2. *$K_W(t,\cdot): T \to \mathbb{R}$ is bi-Lipschitz for γ-a.e. $t \in T$, with*

$$\int_{t\in T} \mathrm{Lip}_{\mathrm{lower}}(K_W(t,\cdot))^2 d\gamma(t) > 0$$

3. *The Karhunen-Loeve transform is equivalent to nondegenerate Gaussian measures (see Proposition 5).*

The core assumptions of our argument are square-integrability of the stochastic process W, (bi-)Lipschitz-ness of K_W, and that the index set of the process is a finite-dimensional manifold. At this level of generality, we are able to prove the following result:

Theorem 1 (Main Result)**.** *Suppose that Assumption 1 is satisfied. Consider the following two conditions:*

(i) For each $k \in \{1,\ldots,n\}$,

$$\mathcal{H}^{k-1}\left(\{t \in T : \{DJ_W(t) \text{ is contained in a } k\text{-dimensional open half-space}\}\right) = 0$$

(ii) $\zeta_W \ll \gamma$, i.e. the argmax measure is absolutely continuous.

Then, (i) $\implies$ (ii). If Assumption 2 is also satisfied then the converse holds, i.e. (ii) $\implies$ (i).

As a result of our analysis, we also obtain a statement about uniqueness of the argmax.

Corollary 1. *Suppose that W is square-integrable, centered, and has a continuous covariance kernel. Then, $Z_W \equiv \arg\max_{t\in T} W_t$ is single-valued $\mathbb{P}$-almost surely.*

2.3 Outline of Proof of Theorem 1

The proof is split into three main steps, and if we are careful we will be able to reverse each step in order to prove the converse:

1. Reduce condition (ii) in Theorem 1 to a geometric statement via a reproducing kernel Hilbert space (RKHS).

2. Use geometric measure theory and variational analysis to find a geometric condition in an abstract Hilbert space that is equivalent to the reduced version of (ii).

3. Translate the abstract geometric condition to a condition on the stochastic process, obtaining (i).

The above are performed in the next three sections and result in Proposition 1, Theorem 3, and Corollary 3 respectively, and are then combined in Section 6.

3 Reduction to Metric Projection

Note that K_W continuous and W square-integrable imply $K_W \in L^2(T \times T, \gamma \otimes \gamma)$. Note further that for each $\omega \in \Omega$ we have $\int_{t \in T} |W_t(\omega)|^2 d\gamma(t) < \infty$; overloading notation, we can therefore express $W : \Omega \to L^2(T, \gamma)$ sending $\Omega \ni \omega \mapsto [t \mapsto W_t(\omega)] \in L^2(T, \gamma)$. Our first goal is to reduce the operation of taking the argmax of W to something more geometric. To do so, we appeal to the well-known Karhunen-Loeve expansion.

Theorem 2 (Theorem 1.2.1(1) in [3])**.** *Suppose that W is a square-integrable, centered stochastic process indexed by γ with a covariance function $K_W : T \times T \to \mathbb{R}$ that is continuous and square-integrable on $T \times T$. Let $(\psi_j)_j \subseteq L^2(T, \gamma)$ and $(\lambda_j)_j \subseteq (0, \infty)$ be the orthonormal basis of eigenfunctions and eigenvalues associated with the compact, self-adjoint, positive integral operator F_W sending $L^2(T, \gamma) \ni f \mapsto \int_T K_W(\cdot, t) f(t) d\gamma(t) \in L^2(T, \gamma)$, and define the random variables $\xi_j : \Omega \to \mathbb{R}$ via*

$$\xi_j(\omega) := \langle W(\omega), \psi_j \rangle_{L^2(T,\gamma)} \quad (\forall \omega \in \Omega)$$

Then, these random variables $(\xi_j)_j$ are centered, uncorrelated, have variance λ_j, and satisfy

$$W(\omega) = \sum_{j=1}^{\infty} \xi_j(\omega) \psi_j(t) \quad (\forall \omega \in \Omega),$$

where the above expression converges uniformly in j in the sense that

$$\sum_{j=1}^{N} \xi_j \psi_j \to W \text{ in } L^2(\Omega \times T, \mathbb{P} \otimes \gamma)$$

uniformly w.r.t. j as $N \to \infty$. Furthermore, if W is a Gaussian process then the random variables $(\xi_j)_j$ are independent Gaussians and the convergence can be upgraded so that for each $t \in T$ the following holds $\mathbb{P}$-a.s.:

$$W_t = \sum_{j=1}^{\infty} \xi_j \psi_j(t)$$

This representation is very nice, because for square-integrable stochastic processes it allows us to evaluate W_t as an inner product in a Hilbert space between a random vector and a fixed embedding of t. To be precise, we know that since K is bounded and compactly-supported then $\sum_j |\psi_j(t)|^2 < \infty$ for each t, and so we can express define a "feature map" $\Phi : T \to X$ to the Hilbert space $X := \ell^2(\mathbb{R})$ via

$$\Phi(t) := (\ldots, \psi_j(t), \psi_{j+1}(t), \ldots)$$

Similarly, for each $\omega \in \Omega$ we note that $\sum_j |\xi_j(\omega)|^2 < \infty$ by construction, and so we can consider a "random feature" $\xi : \Omega \to X$ via

$$\xi(\omega) := (\ldots, \xi_j(\omega), \xi_{j+1}(\omega), \ldots)$$

The Karhunen-Loeve representation states that for all $t \in T$,

$$W_t = \langle \xi, \Phi(t) \rangle_X \quad \mathbb{P}\text{-a.s.}$$

The argmax random variable Z_W can then be expressed as

$$Z_W(\omega) = \arg\max_{t \in T} \langle \xi(\omega), \Phi(t) \rangle_X \quad (\text{for } \mathbb{P}\text{-a.e. } \omega \in \Omega) \tag{$\star$}$$

Remark 2. It is **crucial** that the Karhunen-Loeve transform holds for all $t \in T$ so that for $\mathbb{P}$-a.e. realization of the randomness $\omega \in \Omega$, the relationship in $(\star)$ is exact w.r.t. T (and not up to a γ-null set, for example). This means that we can study absolute continuity of ζ_W w.r.t. γ via absolute continuity of $\mathrm{Law}(\arg\max \langle \xi, \Phi(\cdot) \rangle_X)$ w.r.t. γ.

Lemma 1. *Under Assumption 1, the feature map $\Phi : T \to X$ is Lipschitz. If we also assume Assumption 2, then Φ is bi-Lipschitz.*

Proof. We have that

$$\begin{aligned}
\|\Phi(s) - \Phi(t)\|_X^2 &= \sum_{j \in \mathbb{N}} \left| \langle \psi_j, K_W(s, \cdot) - K_W(t, \cdot) \rangle_H \right|^2 \\
&= \sum_{j \in \mathbb{N}} \frac{1}{\lambda_j} \left| \langle \psi_j, K_W(s, \cdot) - K_W(t, \cdot) \rangle_{L^2(T,\gamma)} \right|^2 \\
&= \left\| F_W^{-1/2} \left(K_W(s, \cdot) - K_W(t, \cdot) \right) \right\|_{L^2(T,\gamma)}^2
\end{aligned}$$

where the first line is due to the reproducing property of the reproducing kernel Hilbert space H for the kernel K_W, the second is the definition of the inner product in H, and the third uses the diagonalization of F_W (along with the fact that it is self-adjoint). Since $F_W^{-1/2}$ is an operator bounded from above and from below, we see that there are $c, C \in (0, \infty)$ s.t.

$$c\|K_W(s,\cdot) - K_W(t,\cdot)\|_{L^2(T,\gamma)}^2 \leq \|\Phi(s) - \Phi(t)\|_X^2 \leq C\|K_W(s,\cdot) - K_W(t,\cdot)\|_{L^2(T,\gamma)}^2$$

Now, if $K_W(\cdot, q)$ is Lipschitz for γ-a.e. q with square-integrable Lipschitz constants, then

$$\|K_W(s, \cdot) - K_W(t, \cdot)\|_{L^2(T,\gamma)}^2 \lesssim \|s - t\|^2,$$

and so Φ is Lipschitz. Under Assumption 2, we also get that

$$\|s - t\|^2 \lesssim \|K_W(s, \cdot) - K_W(t, \cdot)\|_{L^2(T,\gamma)}^2$$

So, Φ is bi-Lipschitz. □

Given Lemma 1, we see that $E := \Phi(T) \subseteq X$ is a $\dim_{\mathcal{H}}(T) =: n$-dimensional closed subset of the separable, infinite-dimensional Hilbert space X. Furthermore, Φ carries γ to a measure that is absolutely continuous w.r.t. $\mathcal{H}^n$ on X (under Assumption 2 even $\Phi \circ \gamma \asymp \mathcal{H}^n$ is true). We have reduced the problem in the following way:

Proposition 1. *Suppose that Assumption 1 holds. If*

$$\mathrm{Law}\left(\arg\max_{y\in E}\langle\xi, y\rangle_X\right) \ll \mathcal{H}^n \llcorner E \tag{†}$$

holds over X, then (ii) from Theorem 1 holds. Under Assumption 2, (ii) $\Longrightarrow$ (†) as well.

Proof. Suppose that (†) holds. Then, for all $A \subseteq E$ with $\mathcal{H}^n(A) = 0$, we have that

$$\mathbb{P}\left[\left(\arg\max_{y\in E}\langle\xi, y\rangle_X\right) \cap A \neq \emptyset\right] = 0$$

So, for all $B \subseteq T$ with $\gamma(B) = 0$, we see that $\mathcal{H}^n(B) = 0$ and so $\mathcal{H}^n(\Phi(B)) = 0$ by Assumption 1(2,3). Thus,

$$\mathbb{P}\left[\left(\arg\max_{t\in T}\langle\xi, \Phi(t)\rangle_X\right) \cap B \neq \emptyset\right] = 0 \implies \zeta_W(B) = 0$$

and so (ii) holds. The converse direction follows under the extra assumptions that $K_W(t, \cdot)$ is bi-Lipschitz for γ-a.e. $t \in T$ and $\gamma \ll \mathcal{H}^n$ on T, and are used solely in order to show that $\mathcal{H}^n(\Phi(B)) = 0 \implies \gamma(B) = 0$. □

4 Geometric Analysis

We now seek to find necessary and sufficient conditions on E under which (†) holds, and we will proceed in the general geometric case of $E \subseteq X$ an arbitrary closed subset. Note that for all $y \in E$,

$$2\langle\xi, y\rangle = \|\xi\|^2 + \|y\|^2 - \|\xi - y\|^2,$$

and so maximizing $\langle\xi, y\rangle$ w.r.t. y is the same as minimizing $\|\xi - y\|^2 - \|y\|^2$ w.r.t. y, i.e.

$$(\dagger) \iff \mathrm{Law}\left(\arg\min_{y\in E}(\|y-\xi\|^2 - \|y\|^2)\right) \ll \mathcal{H}^n \llcorner E$$

The above is almost like a metric projection onto E, but regularized to prefer the larger-norm points of the set E. For full generality, we consider the setting where this regularization can take other values, parameterized by α. Let $\mu := \mathrm{Law}(\xi)$ denote the distribution measure of ξ on X; we defer it until Proposition 5, but under Assumption 1 we may suppose without loss of generality that μ is a nondegenerate Gaussian measure on X.

Definition 6 (Metric Projection & Projection Measure)**.** *Let X be a separable Hilbert space and let $\alpha \in (-\infty, 1]$. Let $P_{E,\alpha} : X \to 2^E$ be the* α-skewed metric projection *map to a closed set $E \subseteq X$ that sends*

$$x \mapsto \left\{y \in E : y \in \arg\min_{z\in E}(\|z - x\|^2 - \alpha\|z\|^2)\right\}$$

Consider the preimage map $P_{E,\alpha}^{-1} : E \to 2^X$ *that maps* $E \ni y \mapsto \{x \in X : y \in P_{E,\alpha}(x)\}$*. For any set* $F \subseteq E$*, extend* $P_{E,\alpha}^{-1}(F) := \bigcup_{y \in F} P_{E,\alpha}^{-1}(y)$ *in the obvious way. We define the* projection measure *induced on* E *by* μ *to be* $\eta_{E,\alpha} := (P_{E,\alpha})_{\#}\mu$*, i.e.*

$$\eta_{E,\alpha}(A) = \mu(P_{E,\alpha}^{-1}(A)) \quad \forall A \subseteq E \text{ measurable.}$$

Note that $\operatorname{spt}(\eta_{E,\alpha}) \supseteq E \cap \operatorname{spt}(\mu)$*.*

In this language, we already have $(\dagger) \iff \eta_{E,1} \ll \mathcal{H}^n \llcorner E$[2]. The goal of this section is therefore to find a geometric condition on E which is equivalent to the condition that $\mathcal{H}^n(A) = 0 \implies \mu(P_{E,\alpha}^{-1}(A)) = 0$ for all $A \subseteq E$. First, a technical lemma that we will use later to handle the $\alpha = 1$ case.

Lemma 2. *If* $x \in X$ *and* $(\alpha_j)_j \subseteq (-\infty, 1)$ *is such that* $\alpha_j \to \alpha$*, then*

$$P_{E,\alpha}(x) \subseteq \bigcap_{m \in \mathbb{N}} \bigcup_{j \geq m} P_{E,\alpha_j}(x)$$

Proof. {evan: **do the proof. it might just be to use the uniform continuity of** $\|z - x\|^2 - \alpha\|z\|^2$ **as a function of** α} □

4.1 Geometric Preliminaries

We will be investigating the behavior of $P_{E,\alpha}^{-1}(\cdot)$ in terms of the tangent structure of E via tools from variational analysis. Most of methods were taken from the standard variational analysis textbook by Rockafellar and Wets [4]. The structure we investigate is (naturally) the tangent cone and its polar, the normal cone.

Definition 7 (Tangent Cone, Definition 6.1 [4])**.** *Let* $x \in E \subseteq X$*, where* X *is a Hilbert space. We define the* tangent cone *to* E *at* $y \in E$ *via*

$$T_E(y) := \{0\} \cup \left\{ z \in E : \exists (y_k)_k \subseteq E,\ (a_k)_k \subseteq (0, \infty) \text{ s.t. } y_k \to y \text{ and } \frac{y_k - y}{a_k} \to z \right\}$$

The tangent cone is the set of directions from which one can approach y within E. We see that $T_E(y)$ is always a closed set and it is also a cone (i.e. $v \in T_E(y) \implies \lambda v \in T_E(y)\ \forall \lambda > 0$); however, $T_E(y)$ need not be convex. An important construction in the geometry of cones is that of the polar cone.

Definition 8 (Normal Cone, see Proposition 6.5 in [4])**.** *Let* $C \subseteq X$ *be a cone, where* X *is a Hilbert space. We define the* polar cone *to* C *via*

$$C^\circ := \{x \in X : \langle x, v \rangle \leq 0\ \forall v \in C\}$$

We define the regular normal cone *to* E *at* $y \in E$ *to be* $\widehat{N}_E(y) := T_E(y)^\circ$*.*

[2] The unskewed setting ($\alpha = 0$) captures the case of the usual nearest-point metric projection (which may be of independent interest), and so we proceed for general $\alpha \in (-\infty, 1]$.

C° is always a closed, convex cone, and it can be shown that $(C^\circ)^\circ = \overline{\operatorname{co}(C)}$, where $\operatorname{co}(\cdot)$ denotes the convex hull operation. We stop and collect some properties of this tangent structure that are standard in the variational analysis/nonsmooth optimization literature.

Proposition 2. *Let $E \subseteq X$ be a subset of a separable Hilbert space X, and let $y \in E$. Then, the following are true*

(i) If $y \in \operatorname{int}(E)$, then $\widehat{N}_E(y) = \{0\}$.

(ii) If E is an n-dimensional C^1 manifold around x (i.e. for some $\delta > 0$, we can represent $B_\delta(x) \cap E$ as the solution set of $F(z) = 0$, where $F : B_\delta(x) \cap E \to X/\mathbb{R}^n$ is a C^1 mapping with $DF(x)$ a full-rank operator), then

$$T_E(x) = \ker(DF(x)) \quad \text{and} \quad \widehat{N}_E(x) = T_E(x)^\perp$$

(iii) If $f : X \to \mathbb{R}$ is (Frechet-) differentiable and $y \in E$ is locally optimal in the constrained optimization problem $\min_{z\in E} f(z)$, then

$$-\nabla f(y) \in \widehat{N}_E(y)$$

Proof. (i) follows plainly from the fact that $T_E(y) = X$ when y is an interior point, while (ii) is Example 6.8 in [4]. (iii) is Theorem 6.12 in [4], and is a standard tool in constrained optimization. □

Remark 3. (ii) is included above simply to contextualize the definitions of tangent and normal cones that we use. When E is locally-linearly-approximable, its tangent cone is a flat plane and so the regular normal cone is the orthogonal complement of the tangent plane as expected. Furthermore, when E is the graph of a C^1 function, the tangent space is given by the Jacobian of the function. These definitions appear to generalize the notions of geometry that we are familiar with.

Corollary 2. *Let $E \subseteq X$ be a closed subset of a Hilbert space, $y \in E$, and $\alpha \in (-\infty, 1]$. Then,*

$$P_{E,\alpha}^{-1}(y) \subseteq (1-\alpha)y + \widehat{N}_E(y),$$

where the right hand side is to be understood as the set $\{(1-\alpha)y + v : \ v \in \widehat{N}_E(y)\}$.

Proof. Fix an $x \in P_{E,\alpha}^{-1}(y)$. Then, y is a local optimum of the constrained optimization problem

$$\min_{z\in E} \frac{1}{2}\left(\|z-x\|^2 - \alpha\|z\|^2\right) =: \min_{z\in E} f_x(z)$$

where $f_x : X \to \mathbb{R}$. Clearly, the objective function is (Frechet-) differentiable with gradient $\nabla f_x(z) = (z-x) - \alpha z = (1-\alpha)z - x$, and so Proposition 2(iii) guarantees that

$$x + (\alpha - 1)y \in \widehat{N}_E(y) \implies x \in (1-\alpha)y + \widehat{N}_E(y)$$

□

Corollary 2's statement that the metric projection preimage is contained in the normal cone is a strong connection between properties of the metric projection map $P_{E,\alpha}$ and the geometry of E itself, and it is the main tool we take from this subsection. If it also happens that the converse holds (and so the $\subseteq$ above is morally an equality), we have the following condition:

Definition 9 (Normal-Regularity[3])**.** *Let $r \in (0, \infty)$. We say a closed set $E \subseteq X$ is* r-normal-regular *iff for all $y \in E$, for all $v \in \widehat{N}_E(y) \cap B_1(0_X)$, and for all $t \in (0, r)$ one has $y \in P_{E,\alpha}(y + tv)$. Letting $B_r(E) := \{x \in X : \ d(x, E) < r\}$ denote the open neighborhood around E of radius r, this is equivalent to stating that*

$$P^{-1}_{E,\alpha}(y) \cap B_r(E) = \Big((1-\alpha)y + \widehat{N}_E(y)\Big) \cap B_r(E)$$

4.2 Measure Theory

We will phrase things for an appropriate measure-theoretic notion "small" with which to prove continuity. Namely, we prove our geometric results for μ in the class of nondegenerate (i.e. strictly positive covariance operator) Gaussian measures on X, which we will then broaden to the general case $\mu = \mathrm{Law}(\xi)$[4].

Definition 10 (Null Sets)**.** *A measurable set $A \subseteq X$ is* Gauss-null *if $\mu(A) = 0$ for all nondegenerate Gaussian measures μ on X. This condition is equivalent to* Aronszajn-null *and* cube-null *and implies* Haar-null *(see [5]). If X is finite-dimensional, a set is Gauss-null iff it is Lebesgue-null.*

We seek conditions on E for which $S \subseteq E$ with $\mathcal{H}^n(S) = 0 \implies P^{-1}_{E,\alpha}(S)$ is Gauss-null. By Corollary 2, we may bound $P^{-1}_{E,\alpha}(S) \subseteq \bigcup_{y \in S} \Big((1-\alpha)y + \widehat{N}_E(y)\Big)$. Our method will be to cover this by an uncountable (but sufficiently small) union of translates of subspaces of X, and show that that the union of these subspaces is Gauss-null. Measure-theoretically, this (measuring an uncountable union) is a scary operation, but we will succeed if we proceed carefully.

Lemma 3 (Uncountable Union of Affine Planes)**.** *Fix $k \in \mathbb{N}$ and $\alpha \in (-\infty, 1]$. Let $S \subseteq X$ be a subset with $\mathcal{H}^k(S) = 0$. For each $x \in S$, let $V_x \in \mathrm{Gr}(X, k)$ denote a k-dimensional subspace of X. Then,*

$$\bigcup_{x \in S} ((1-\alpha)x + V_x^{\perp}) \text{ is Gauss-null}$$

Proof. We will first reduce this to the case for finite-dimensional X via a finite-dimensional approximation argument for Gaussian measures. project to finite-dim approximation, use coarea formula on product space, profit So, we may take $X \cong \mathbb{R}^{n+k}$. ☐

[3]Note that this condition is more restrictive than the related one of *r-prox-regularity* that is studied in related literature. See Definition 1 and Theorem 1 in [6] for context, as well as the discussion after Example 6.16 in [4] for the explanation of the relationship between regular and proximal normals.

[4]In the case W is a Gaussian process, we saw earlier in Theorem 2 that $\mathrm{Law}(\xi)$ is a nondegenerate Gaussian measure on X anyway.

Lemma 4 (Uncountable Union of Affine Half-Planes)**.** *Fix $k \in \mathbb{N}$ and $\alpha \in (-\infty, 1]$. Let $S \subseteq X$ be a subset with $\mathcal{H}^k(S) > 0$. For every $x \in S$, let H_x be an arbitrary $(k+1)$-dimensional vector half-subspace of X. Then,*

$$\bigcup_{x \in S} ((1-\alpha)x + H_x^\circ) \text{ has nonempty interior in } X$$

Proof. Firstly, we note that if H_x is a $(k+1)$-dimensional half-space (i.e. up to some unitary, the $(k+1)^{th}$ coordinate is nonnegative and everything afterward is 0), then H_x° is a codimension-k half-space (up to the same unitary, it is the set of vectors with the $(k+1)^{th}$ coordinate nonpositive and everything before taking value 0). **DO THIS. Boris insists both of these lemmas are true.** □

4.3 Main Geometric Result

With the dichotomy given by Lemmas 3 and 4, we are able to craft a sufficient and (almost) necessary condition for our desired property to hold:

Definition 11 (m-Pointedness)**.** *For $m \in \mathbb{N} \cup \{0\}$, a set $E \subseteq X$ is said to be* m-pointed *at $y \in E$ iff $\widehat{N}_E(y)^\circ \equiv \overline{\mathrm{co}(T_E(y))}$ does not contain any m-dimensional subspace, where $\mathrm{co}(\cdot)$ denotes the convex hull operation. Let $\mathcal{P}_m \subseteq E$ denote the set of points at which E is m-pointed, and so $\emptyset = \mathcal{P}_0 \subseteq \mathcal{P}_1 \subseteq \ldots$. We define $\widetilde{\mathcal{P}}_k := \mathcal{P}_{k+1} \setminus \mathcal{P}_k$ for $k \in \mathbb{N} \cup \{0\}$, arriving at the* pointedness stratification *of E:*

$$E = \bigcup_{m \in \mathbb{N} \cup \{0\}} \mathcal{P}_m = \bigsqcup_{k \in \mathbb{N} \cup \{0\}} \widetilde{\mathcal{P}}_k$$

Proposition 3. *For a closed set $E \subseteq X$ and a point $y \in E$, the following conditions are equivalent:*

(i) E is m-pointed at y.

(ii) $T_E(y) \setminus \{0\}$ is contained in an open m-dimensional half-space of X.

(iii) {evan: something involving inner products}

Proof. (i $\implies$ ii) We prove this by contrapositive: namely, we wish to show that if the closed convex cone $C := \overline{\mathrm{co}(T_E(y))}$ is such that $C \setminus \{0\}$ is contained in no open m-dimensional half-space of X, then C must contain an m-dimensional subspace. It is clear how the implication (i $\implies$ ii) follows if we can show this. Since C is a convex cone, this amounts to finding m pairwise orthogonal vectors in $C \cap -C$. We do this iteratively: for each step $k \in \{1, \ldots, m\}$ we use that $C \setminus \{0\}$ is not contained in any open k-dimensional half-space in order to select a direction u_k s.t. $u_k \in C \cap -C$ and $\langle u_k, u_j \rangle_X = \delta_{jk}$ for all $j \leq k$. This selection can be done by considering any open k-dimensional half-space and its reflection: in order for $C \setminus \{0\}$ to be contained in neither, it must hold that $(C \cap -C) \setminus \{0\}$ is not contained in the half-space either. So, since $(C \cap -C) \setminus \{0\}$ is not contained in any k-dimensional half-spaces, it must contain a vector $\widetilde{u}_k$

orthogonal to the span $\{u_j\}_{j=1}^{k-1}$. As C is a cone, we can select a unit vector $u_k := \widetilde{u}_k/\|\widetilde{u}_k\| \in C \cap -C$, and so the orthonormal set satisfies $\mathrm{span}\{u_j\}_{j=1}^{k} \subseteq C$. Iterating this procedure until $k = m$, we get that C contains an m-dimensional half-space.

(ii $\implies$ i) If $T_E(y)$ is contained in an m-dimensional half-space, then so is $\overline{\mathrm{co}(T_E(y))}$. Clearly, this means that $\overline{\mathrm{co}(T_E(y))}$ may not contain an m-dimensional subspace, and so E is m-pointed at y. $\square$

Theorem 3 (Geometric Criterion for Absolute Continuity)**.** *Let X be a separable Hilbert space, $E \subseteq X$ closed, and $n \in \mathbb{N}$. Consider the following conditions:*

(i) $\mathcal{H}^k(\widetilde{\mathcal{P}}_k) = 0$ for all $k \in \{0, \dots, n-1\}$.

(ii) $\eta_{E,\alpha} \ll \mathcal{H}^n \llcorner E$ for all nondegenerate Gaussian measures μ and all $\alpha \in (-\infty, 1]$ (i.e. the projection measure on E is a.c. w.r.t. $\mathcal{H}^n \llcorner E$).

Then, (i) $\implies$ (ii) and, if E is r-normal-regular for some $r > 0$, (ii) $\implies$ (i).

Proof. (i) $\implies$ (ii). Suppose that $\mathcal{H}^k(\widetilde{\mathcal{P}}_k) = 0$ for $k \le n-1$. Fix $\alpha \in (-\infty, 1]$ and let $S \subseteq E$ be such that $\mathcal{H}^n(S) = 0$. Stratify

$$S = \bigsqcup_{k \in \mathbb{N} \cup \{0\}} (S \cap \widetilde{\mathcal{P}}_k)$$

Define $\widetilde{T}_k := S \cap \widetilde{\mathcal{P}}_k$ for notation, and so by countable additivity it suffices to show that $P_{E,\alpha}^{-1}(\widetilde{T}_k)$ is Gauss-null for all $k \in \mathbb{N} \cup \{0\}$. By stratification, for all $y \in \widetilde{T}_k$ we know that the closed convex cone $\widehat{N}_E(y)^\circ$ contains no $(k+1)$-dimensional subspace, but *does* contain a k-dimensional subspace (if not, then y would instead belong to $\widetilde{T}_{k-1}$ or below; note that this means $\widetilde{T}_k = \emptyset$ for $k \ge n$ since S cannot contain a subspace of dimension n). For each $y \in \widetilde{T}_k$, let $V_y \subseteq \widehat{N}_E(y)^\circ$ be such a k-dimensional subspace. The monotonicity property of polar cones and Corollary 2 give that

$$V_y \subseteq \widehat{N}_E(y)^\circ \implies \widehat{N}_E(y) \subseteq V_y^\perp \implies P_{E,\alpha}^{-1}(\widetilde{T}_k) \subseteq \bigcup_{y \in \widetilde{T}_k} ((1-\alpha)y + V_y^\perp)$$

Applying the assumption and Lemma 3 completes this direction.

(ii) $\implies$ (i). We will show the contrapositive. Suppose that (i) does not hold, and so there is some $k \in \{0, \dots, n-1\}$ for which $\mathcal{H}^k(\widetilde{\mathcal{P}}_k) > 0$. Select a set $G \subseteq \widetilde{\mathcal{P}}_k$ with $\mathcal{H}^k(G) > 0$ but $\mathcal{H}^n(G) = 0$, which we may do since $k \le n-1$. At each $y \in \widetilde{\mathcal{P}}_k \subseteq \mathcal{P}_{k+1}$, Proposition 3(ii) yields that $T_E(y) \setminus \{0\}$ must be contained in a $(k+1)$-dimensional open half-subspace; call it H_y. Then,

$$\bigcup_{y \in G} ((1-\alpha)y + H_y^\circ) \subseteq \bigcup_{y \in G} ((1-\alpha)y + \widehat{N}_E(y))$$

By Lemma 4, the set $A := \bigcup_{y\in G}((1-\alpha)y + \widehat{N}_E(y))$ has nonempty interior in X. When $\alpha < 1$, we note that $1-\alpha > 0 \implies (1-\alpha)(y+\widehat{N}_E(y)) = (1-\alpha)y+(1-\alpha)\widehat{N}_E(y) = (1-\alpha)y+\widehat{N}_E(y)$ and so $A = (1-\alpha)\bigcup_{y\in G}\left(y+\widehat{N}_E(y)\right)$. However, if $\alpha = 1$ then $A = \bigcup_{y\in G}\widehat{N}_E(y)$. Now, let $B_r(E)$ denote the open neighborhood around E of radius r, where $r > 0$ is such that E is r-normal-regular. Then, the r-normal-regularity condition reads that for all $\alpha \in E$,

$$B_r(E) \cap P_{E,\alpha}^{-1}(G) = B_r(E) \cap \left(\bigcup_{y\in G}\left(y+\widehat{N}_E(y)\right)\right)$$

In the case $\alpha < 1$, the right hand side is none other than $B_r(E) \cap \frac{1}{1-\alpha}A$. Since the intersection is nonempty and $B_r(E)$ is open in X, we see that $B_r(E) \cap P_{E,\alpha}^{-1}(G)$ has nonempty interior, which means that $P_{E,\alpha}^{-1}(G)$ does too. We handle the case $\alpha = 1$ via Lemma 2 and a limiting argument, and so in any case we find that $P_{E,\alpha}^{-1}(G)$ has nonempty interior. Letting μ be a nondegenerate Gaussian measure, we know that μ assign positive measure to open sets. Thus, $\mu(P_{E,\alpha}^{-1}(G)) > 0$. However, we selected G so that $\mathcal{H}^n(G) = 0$, and so the projection measure cannot have a density w.r.t. $\mathcal{H}^n$ over G. □

4.4 *Digression into Stationarity

This subsection assumes a good amount of prerequisite knowledge of geometric measure theory (GMT); for this, we recommend [8] as a classic, comprehensive source and [7] for more detail specifically on regularity of stationary varifolds.

For a certain (n-rectifiable) set $E \subseteq X$, we can instead consider the varifold $\|V\| := \mathcal{H}^n \llcorner E$ with unique tangent planes $\mathcal{H}^n$-a.e.. Note that V will always be an integral varifold, as $\Theta(E,x) = 1$ for $\mathcal{H}^n$-a.e. $x \in E$ is a consequence of rectifiability. We concern ourselves with k-pointedness at the singular points of V, and in particular with this property's relationship to the following fundamental concept of stationarity.

Definition 12 (Stationary Sets, Definition 2.6 in [8])**.** *An n-rectifiable set $E \subseteq X$ is* n-stationary *in X if the first variation vanishes, i.e. if*

$$\int_E \operatorname{div}_E F d\mathcal{H}^n = 0$$

for all C^1 vector fields F on X with compact support.

Such sets generalize the classical notion of "area-minimizing surfaces", as this is equivalent to the mean curvature vanishing[5]. If E is the graph of a function u with enough regularity, E is stationary iff u is a solution to the minimal surface PDE. For such sets, some extra regularity is established for free, which we can

[5]Many results in GMT that require stationary sets can be generalized directly to sets with bounded (instead of vanishing) mean curvature, see Theorem 3.14 in [8].

relate to our setting. Eventually, we would like to show that stationarity $\implies$ condition (i) from Theorem 3.

As a first pass, we observe that the condition (i) in Theorem 3 is phrased in terms of a stratification by pointedness. If $x \in \widetilde{\mathcal{P}}_k$ then $\overline{\text{co}(T_E(x))}$ cannot contain a $(k+1)$-dimensional subspace, and so $T_E(x)$ certainly cannot have a $(k+1)$-dimensional spine, which means $x \in \mathcal{S}_k$ and so $\widetilde{\mathcal{P}}_k \subseteq \mathcal{S}_k$ (this is the relationship to Almgren's stratification by dimension of spine of the tangent cone, see Definition 1.6 in [7]). By Theorem 1.7 in [7], if E is stationary (area-minimizing) we automatically have that $\dim_{\mathcal{H}}(\widetilde{\mathcal{P}}_k) \leq k$. Thus, for stationary E we don't yet have a way to guarantee that $\mathcal{H}^k(\widetilde{\mathcal{P}}_k) = 0$, but we already know $\mathcal{H}^{k+\delta}(\widetilde{\mathcal{P}}_k) = 0$ for all $\delta > 0$[6]. Close, but not quite there.

The intuition underlying the previous result is that stationary sets cannot "lie too much on one side" at an infinitesimal scale (since one could reduce the volume by moving in that direction, and so it cannot be a stationary point for volume). This intuition also surfaces through "maximum principles" such as the following, proved by Brian White:

Theorem 4 (Maximum Principle for Minimal Varifolds, Theorem 1 in [11])**.** *Let N be a smooth Riemannian manifold with boundary, and let p be a point in ∂N at which N is strongly m-convex (defined in [11]). Then, p is not contained in the support of any m-dimensional varifold in N that minimizes area to first order in N. Indeed, there is an $\varepsilon > 0$ such that*

$$d(p, \operatorname{spt} \|V\|) \geq \varepsilon$$

for all such varifolds V.

The ambition is for V to represent the (affine) tangent cone at p to our stationary varifold $\mathcal{H}^n \llcorner E$ and to have N be the m-dimensional affine half-space containing V, and so V minimizes area to first order in N. If we can show that m-dimensional half-spaces are "strongly m-convex" along their boundary, then Theorem 4 tells us that $\mathcal{P}_m = \emptyset$ for all $m \leq n$ and so condition (i) in Theorem 1 holds immediately. We implement this program below:

Lemma 5. *If $H \subseteq X$ is an open m-dimensional half-space, then $\overline{H}$ is m-strongly convex on ∂H.*

Proof. H is locally an m-plane everywhere in its interior, which means it is m-strongly convex. {evan: please elaborate evan, im p sure this is false} ☐

Proposition 4 (Continuity of Projection for Stationary Sets)**.** *Suppose that $E \subseteq X$ is a closed, stationary, n-rectifiable set. Then, condition (i) in Theorem 1 holds for E and so the projection measure on E is a.c. w.r.t. $\mathcal{H}^n \llcorner E$.*

[6] Furthermore, stationarity implies apriori that $\widetilde{\mathcal{P}}_k$ is k-rectifiable by a result of Naber-Valtorta cited in the aforementioned notes. Of course, if condition (i) holds then $\mathcal{H}^k(\widetilde{\mathcal{P}}_k) = 0 \implies \widetilde{\mathcal{P}}_k$ is k-rectifiable trivially, but this result become useful while proving condition (i).

Proof. Suppose by way of contradiction that $y \in \mathcal{P}_m$ for some $m \leq n$. Then, by Proposition 3 we know that $T_E(y) \setminus \{0\}$ is contained in an open m-dimensional half-space whose closure we call N. By Lemma 5, N is m-strongly convex at 0. The varifold $C := T_V(y)$ minimizes area to first order in N by stationarity of E, which means we can apply Theorem 4 to get that $0 \notin C \implies p \notin E$. This is a clear contradiction, which means that $\mathcal{P}_m = \emptyset$ for $m \leq n$. Therefore, $\widetilde{\mathcal{P}}_k = \emptyset$ for $k \leq n-1$, yielding condition (i) in Theorem 1. □

5 Returning to Stochastic Processes

To complete the chain of logic for the proof of Theorem 1, we re-impose the extra structure that $E = \Phi(T)$ for a Lipschitz map $\Phi : T \to X$.

5.1 Theorem 1(i) $\iff$ Theorem 3(i)

We shall relate the tangent spaces $T_{\Phi(T)}(\Phi(t))$ to the Gateaux differential $D\Phi(t) \subseteq X$, which we then relate to the directional derivatives of the covariance diagonal J_W. Note that both $Df(t)$ and $\mathrm{dom}(Df)(t)$ are closed cones, but need not be convex.

Lemma 6. *Let $f : U \to V$ be a map between closed subsets of Hilbert spaces. Then,*

$$Df(t) \subseteq T_{f(U)}(f(t))$$

and the reverse inclusion also holds if f has Lipschitz inverse and U is finite-dimensional. In such a setting, $f(U)$ is k-pointed at $f(t) \in X$ if and only if $Df(t)$ is k-pointed at 0.

Proof. ($\subseteq$) Suppose first that $v = f'(t,u)$ exists for some $u \in U$. Then, letting δ approach 0 along $a_k := \frac{1}{k}$ and setting $v_k := f(t+a_k u)$, the directional derivative condition reads

$$v = \lim_{k\to\infty} \frac{v_k - f(t)}{a_k}$$

for $(a_k)_k \subseteq (0,\infty)$ and $(v_k)_k \subseteq f(U)$. Also, by continuity of f and boundedness of T, $v_k \to f(t)$. So, $v \in T_{f(U)}(f(t))$ by definition.

($\supseteq$) Suppose now that f has Lipschitz inverse and $v \in T_{f(U)}(f(t))$ is nonzero (as long as $0 \in U$, the result holds trivially if $v = 0$). Then, there is a sequence $(v_k)_k \subseteq f(U)$ and $(a_k)_k \subseteq (0,\infty)$ s.t. $v_k \to f(t)$ and $\frac{v_k - f(t)}{a_k} \to v$. In other words, there is a sequence $(t_k)_k \subseteq T$ and $(a_k)_k \subseteq (0,\infty)$ s.t. $f(t_k) \to f(t)$ and $\frac{f(t_k)-f(t)}{a_k} \to v$. For each k, define $u_k := \frac{1}{a_k}(t_k - t)$ (so that $t_k = t + a_k u_k$), and so

$$\frac{f(t + a_k u_k) - f(t)}{a_k} \to v$$

Since f has Lipschitz inverse,

$$\|u_k\| = \left\| \frac{t_k - t}{a_k} \right\| \leq \left\| \frac{f(t_k) - f(t)}{a_k} \right\| \to \|v\| < \infty$$

So, by finite-dimensionality of U, up to subsequences we have $u_k \to u \in U$. We note that $f(t + a_k u)/a_k$ and $f(t + a_k u_k)/a_k$ converge to the same limit by continuity of f and the fact $u_k \to u$. Therefore,

$$\frac{f(t + a_k u) - f(t)}{a_k} \to v \implies v = f'(t, u) \implies v \in Df(t)$$

(The rest) The first inclusion tells us the following: if $\overline{\text{co}(T_{f(T)}(f(t)))}$ does not contain any k-dimensional vector subspaces, then neither does $\overline{\text{co}(Df(t))}$. Restating this result, if $f(T)$ is k-pointed at $f(t)$ then $Df(t)$ is k-pointed at 0 (here we use that $T_{Df(t)}(0) = Df(t)$ since $Df(t)$ is a cone). The other direction follows from the fact that if $\overline{\text{co}(T_{f(T)}(f(t)))}$ does contain a k-dimensional subspace, then so does $\overline{\text{co}(Df(t))}$, and so $Df(t)$ is not k-pointed at 0 either. ☐

Corollary 3. *Let $\Phi : T \to X$ be the spectral feature map for the covariance kernel of a square-integrable, centered stochastic process W with covariance kernel K_W. Then, $\Phi(T)$ is k-pointed at $\Phi(t)$ if (and, under Assumption 2, only if) $DJ_W(t)$ is contained in a k-dimensional open half-space.*

Proof. The above lemma yields that $\Phi(T)$ is k-pointed at $\Phi(t)$ if and only if $D\Phi(t)$ is k-pointed at 0. To relate this to pointedness of $DJ_W(t)$, we simply note that for $u, v \in \text{dom}(D\Phi)(t)$, {evan: **FINISH!** the vibe is $\langle \Phi'(t, u), \Phi'(t, v) \rangle_X = K'_W((t, t); (u, v))$} ☐

5.2 Wrapping Up

The only result we have used for which extra assumptions are needed to show the converse direction is Theorem 3, where we required that E was r-normal-regular for some $r > 0$. We handle this now.

Lemma 7. *Suppose that $E = \Phi(T) \subseteq X$ for a bi-Lipschitz map $\Phi : T \to X$, where T is a finite-dimensional vector space and X is a Hilbert space. Then, E is r-normal-regular for some $r > 0$.*

Proof. {evan: **prove this too**} ☐

There is one final task that prevents us from completing the proof. In Section 4 we showed absolute continuity of the (skewed) metric projection of nondegenerate Gaussian measures μ, yet in application we would like μ to describe the distribution of the random variable $\xi : \Omega \to X$. So, it is left to verify the correct continuity of the Karhunen-Loeve transform. For the forward direction, in order to show condition (ii) of Theorem 1 we want to show $\mathbb{P}(\{\xi \in A\}) = 0$, but we have $\nu(A) = 0$ for nondegenerate Gaussians ν. So, we require that $\text{Law}(\xi) \equiv \mu \ll \nu$ for some nondegenerate Gaussian ν in order to complete the proof of the forward direction. For the converse, Assumption 2 is constructed exactly so that $\mathbb{P}(\{\xi \in A\}) = 0 \implies \nu(A) = 0$, which is enough to relate the result of Theorem 3 to the bigger picture.

Proposition 5. *Let μ be the distribution of the random variable ξ from Theorem 2. Then, under Assumption 1, $\mu \ll \nu$ for some nondegenerate Gaussian measure ν on X.*

Proof. We first note that the claim holds if for all $j \in \mathbb{N}$, the random variable ξ_j is equivalent to a 1-dimensional standard Gaussian (which is, in turn, equivalent to $\mathcal{L}^1$). So, we wish to show that for all $j \in \mathbb{N}$ the random variable $\xi_j : \Omega \to \mathbb{R}$ given by

$$\xi_j(\omega) := \langle W(\omega), \psi_j \rangle_{L^2(T,\gamma)} = \int_{t \in T} W_t(\omega)\psi_j(t) d\gamma(t)$$

is absolutely continuous w.r.t. Lebesgue on $\mathbb{R}$. We know that $\|\xi(\omega)\|_X^2 = \|W(\omega)\|_{L^2(T,\gamma)}^2$ by orthonormality of the basis $(\psi_j)_j$, and so

$$\infty > \sum_{j \in \mathbb{N}} \xi_j^2(\omega) = \sum_{j \in \mathbb{N}} \xi_j(\omega) \int_{s \in T} W_s(\omega) d\gamma(s) = \sum_{j \in \mathbb{N}} \int_{T^2} W_t(\omega) W_s(\omega) \psi_j(t) d\gamma^2((s,t))$$

Thus,

$$\begin{aligned}
\sum_{j \in \mathbb{N}} \mathbb{E}\left[\xi_j^2/\lambda_j\right] &= \frac{1}{\lambda_j} \sum_{j \in \mathbb{N}} \int_{T^2} \mathbb{E}\left[W_t W_s\right] \psi_j(t) d\gamma^2((s,t)) \\
&= \sum_{j \in \mathbb{N}} \frac{1}{\lambda_j} \int_T \int_T K(t,s) \psi_j(t) d\gamma(t) d\gamma(s) \\
&= \sum_{j \in \mathbb{N}} \frac{1}{\lambda_j} \int_T \lambda_j \psi_j(s) d\gamma(s) \\
&= \sum_{j \in \mathbb{N}} \int_T \psi_j(s) d\gamma(s)
\end{aligned}$$

{evan: finish proof that $W(\omega) \in L^2(T,\gamma)$ is actually an element of the RKHS H} Therefore, the reproducing property holds, and so

$$\begin{aligned}
W_t(\omega) &= \langle W(\omega), K(t,\cdot) \rangle_H \\
&= \sum_{j \in \mathbb{N}} \frac{1}{\lambda_j} \langle W(\omega), \psi_j \rangle_{L^2(T,\gamma)} \langle \psi_j, K(t,\cdot) \rangle_{L^2(T,\gamma)} \\
&= \sum_{j \in \mathbb{N}} \xi_j(\omega) \psi_j(t) = \langle \xi(\omega), \Phi(t) \rangle_X
\end{aligned}$$

as expected. {evan: you know what to do} □

6 Proof of Theorem 1

The proof is a straightforward assembly of the results from the previous sections.

Proof of Theorem 1. (i $\implies$ ii) Suppose first that Assumption 1 and condition (i) hold. By Corollary 3, for each $k \in \{1, \ldots, n\}$, the set of points $\Phi(t)$ at which $\Phi(T)$ is k-pointed is $\mathcal{H}^{k-1}$-null. So, by Theorem 3, $\eta_{\Phi(T),1} \ll \mathcal{H}^n \llcorner \Phi(T)$, and we have therefore shown condition (†). So, by Proposition 1, we get condition (ii).

(ii $\implies$ i) Suppose now that Assumption 2 and condition (ii) hold. These conditions satisfy the converse direction of Proposition 1, and so (†) holds. Since $\Phi(T)$ is r-normal-regular for some $r > 0$ by Lemma 7, we can apply the converse of Theorem 3 to see that for each $k \in \{1, \ldots, n\}$, the set of points $\Phi(t)$ at which $\Phi(T)$ is k-pointed is $\mathcal{H}^{k-1}$-null. (i) follows from Corollary 3. □

7 Proof of Corollary 1

Let $\phi_E : X \to \mathbb{R}$ be the functional given by

$$\phi_E(x) := \frac{1}{2} d^2(x, E)$$

Combining results from Asplund, Fitzpatrick, and Aronszajn, we get the following properties:

Lemma 8 (Regularity of $P_{E,0}$ and ϕ_E)**.** *Let $E \subseteq X$ be closed. There is a dense G_δ set $U \subseteq X$ such that $X \setminus U$ is Gauss-null and for all $x \in U$ the following properties hold:*

(i) $P_{E,0}(x)$ is a singleton.

(ii) ϕ_E is Gateaux differentiable at x with derivative (weakly) equal to $\phi'_E(x) = x - P_{E,0}(x)$.

Proof. Asplund's paper and the chain rule give that ϕ_E is Gateaux differentiable on a dense G_δ set $U \subseteq X$ (this can be upgraded to Frechet differentiability, but the next point cannot). Since ϕ_E is locally-Lipschitz, Arnoszajn's theorem on Gateaux differentiation of Lipschitz functions says that such a condition fails on an Aronszajn-null (equivalently, Gauss-null) set, and so $X \setminus U$ is Gauss-null[7]. Gateaux-differentiability $\implies$ (i) is stated as Theorem 2.3(a) in [10]. By Lemma 2.2(a) in [10] and the chain rule, even $\phi'_E(x) = x - P_{E,0}(x)$ is true weakly on U. □

This equips us with enough tools to prove Corollary 1, which claimed that if W is square-integrable (and centered) with continuous kernel then the argmax is unique $\mathbb{P}$-a.s..

Proof of Corollary 1. As we saw in the reduction in Section 3, the set-valued random variable Z_W is single-valued if and only if $\arg\min_{t \in T}(\|\Phi(t) - \xi\|^2 - \|\Phi(t)\|^2)$ is. This in turn happens if and only if $P_{E,1}(\xi)$ is a singleton, and we

[7]See the discussion in Section 3 of this paper.

therefore aim to use Lemma 8 to show that $P_{E,\alpha}$ is a singleton on U for $\alpha \neq 0$ as well. Let $x \in U \setminus \{0\}$ be arbitrary, and suppose that $w, z \in P_{E,\alpha}(x)$. The goal is to show that $w = z$, as this will show that $P_{E,\alpha}$ is unique at x and therefore that $Z_W(\omega) = P_{E,1}(\xi(\omega))$ is unique at $\xi(\omega)$ for $\mathbb{P}$-a.e. $\omega \in \Omega$. By Corollary 2,

$$x \in ((1-\alpha)w + \widehat{N}_E(w)) \cap ((1-\alpha)z + \widehat{N}_E(z))$$

For $\alpha < 1$, the cone structure yields

$$((1-\alpha)w + \widehat{N}_E(w)) \cap ((1-\alpha)z + \widehat{N}_E(z)) = (1-\alpha)\left((w + \widehat{N}_E(w)) \cap (z + \widehat{N}_E(z))\right)$$

This would imply {evan: for $\alpha \notin \{0,1\}$ we need r-normal-regularity, see Lemma 7)} that $x/(1-\alpha) \in P_{E,0}^{-1}(w) \cap P_{E,0}^{-1}(z) \implies z, w \in P_{E,0}(x/(1-\alpha))$. Thus, if $x/(1-\alpha) \in U$ then $z = w$ by Lemma 8(i). Since $X \setminus U$ is Gauss-null, the intersection of U with any 1-dimensional line must have full $\mathcal{H}^1$ measure; applying this with the line $\operatorname{span}\{x\}$ gives that $z = w$ for Lebesgue a.e. $\alpha < 1$.

We may therefore select $(\alpha_j)_j$ such that $\alpha_j \to 1$ and $P_{E,\alpha_j}(x)$ is a singleton. By Lemma 2, it must be that $P_{E,1}(x)$ is the singleton containing the limit

$$\lim_{j\to\infty} P_{E,\alpha_j}(x)$$

if it exists in X. Since this holds for all $x \in U \setminus \{0\}$ and $X \setminus U$ is Gauss-null, the argmax (if it exists) will be unique $\mathbb{P}$-a.s.. □

8 Conclusion

At a high level, we have shown that the argmax measure for a broad class of stochastic processes with finite-dimensional index sets is a.c. over the index set if and (under Assumption 2) only if the Gateaux differential of the covariance kernel is not too pointed not too often. The proof applies a reduction from the argmax of a stochastic process to a skewed metric projection in an abstract Hilbert space, from which abstract geometric conditions for absolute continuity are derived and translated to conditions on the stochastic process itself. We hope that Theorem 1 can be useful to those studying the argmax of a stochastic process, and perhaps Theorem 3, the geometric perspective, and the connection to stationarity are of independent interest.

References

[1] link

[2] link

[3] link

[4] this link

[5] da linky

[6] linkster

[7] minty link

[8] leon link

[9] allard link

[10] Fitzpatrick's paper

[11] Brian White crushed it