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.

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 :)

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

- Fix the off-by-one in the converse; otherwise, (ii) $\implies$ (i) doesn't hold as written in the main theorem.
- Figure out dot-product formulation of k-dimensional open half space-containment property, and prove equivalence of all the formulations of pointedness. Also connect it in Corollary 2.
- Show the two measure theory lemmas that Boris said are "just the Cameron-Martin Theorem".
- Show the second $\subseteq$ relating the tangent cone of $f(U)$ at $f(t)$ to $Df(t)$.
- Prove the $\alpha = 1$ case (maybe in a limiting way) in the proof of Theorem 3.

*Princeton University

- Show r-normal-regularity of $f(U)$ for Lipschitz f.
- Figure out $\mathrm{Law}(\xi)$ and learn about the continuity of the Karhunen-Loeve transform.
- Prove K_W Lipschitz $\implies$ Φ Lipschitz.
- Check if first order optimality condition holds in infinite dimensions.
- Figure out whether square-integrability needs to extend over index set as well.

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. *If $\mathbb{E}|W_t|^2 < \infty$ for all t, we say that W is* square-integrable *{evan: might need square integrability over index set as well, i.e. $W \in L^2(\Omega \times T, \mathbb{P} \otimes \gamma)$}. W is called a* Gaussian process *if $(W_t)_{t \in S}$ is distributed as a Gaussian random vector for any finite $S \subseteq T$.*

Definition 2 (Argmax of Stochastic Process)**.** *Let W be a stochastic process indexed by γ. Define the 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 and compact subset of a Hilbert space. First, three quick definitions in order to state the main result.

[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 $\mathbb{E}\|Z\|^2 < \infty$.

Definition 3 (Covariance Function)**.** *For a square-integrable stochastic process* W*, define the* covariance function $K_W : T \times T \to \mathbb{R}$ *sending* $(s,t) \mapsto \mathbb{E}[W_s W_t] = \langle W_s, W_t \rangle_{L^2(T,\gamma)}$*. 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*

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

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

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

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 Result

Assumption 1. *Let* W *be a stochastic process indexed by* γ *with diagonal covariance function* $J_W : T \to \mathbb{R}$*. 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.* T *is finite-dimensional).*

3. $J_W : T \to \mathbb{R}$ *is Lipschitz.*

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) $\zeta_W \ll \gamma$.

Then, (i) $\implies$ *(ii) always and (ii)* $\implies$ *(i) under the additional assumptions that* $\gamma \ll \mathcal{H}^n \llcorner T$ *and* J_W *is bi-Lipschitz.*

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 2 respectively, and are then combined in Section 6.

3 Reduction to Metric Projection

Let $\mathcal{H}$ denote $L^2(T, \gamma)$ for convenience, and note that K_W continuous and T compact 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 $W : \Omega \to \mathcal{H}$ sending $\Omega \ni \omega \mapsto [t \mapsto W_t(\omega)] \in L^2(T, \gamma) = \mathcal{H}$. 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. Let $(\psi_j)_j \subseteq \mathcal{H}$ and $(\lambda_j)_j \subseteq [0, \infty)$ be the eigenfunctions and eigenvalues associated with the compact, self-adjoint, positive integral operator F_W sending $\mathcal{H} \ni f(\cdot) \mapsto \int_T K_W(\cdot, t) f(t) d\gamma(t) \in \mathcal{H}$, and define the random variables $\xi_j : \Omega \to \mathbb{R}$ via*

$$\xi_j(\omega) := \langle W(\omega), \psi_j \rangle_{\mathcal{H}} \quad (\forall \omega \in \Omega)$$

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

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

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 exactly 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

$$Z_W \stackrel{d}{=} \arg\max_{t \in T} \langle \xi, \Phi(t) \rangle_X \tag{$\star$}$$

Remark 2. It is **crucial** that the Karhunen-Loeve transform holds for all $t \in T$ so that for each realization of the randomness $\omega \in \Omega$, the relationship in ($\star$) is exact w.r.t. T. This means that we can study absolute continuity of ζ_W w.r.t. γ via absolute continuity of $\text{Law}(\arg\max \langle \xi, \Phi(\cdot) \rangle_X)$ w.r.t. γ.

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

Proof. {evan: **PROVE THIS**} $\square$

Given Lemma 1, we see that $E := \Phi(T) \subseteq X$ is a $\dim_{\mathcal{H}}(T) =: n$-dimensional compact 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 (if J_W is bi-Lipschitz then 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. If we also assume that J_W is bi-Lipschitz and $\mathcal{H}^n \llcorner T \asymp \gamma$ then (ii) $\implies$ (†).

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,

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

So, (ii) holds. The converse direction follows under the extra assumptions that J_W is bi-Lipschitz 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. Let $\mu := \mathrm{Law}(\xi)$ denote the distribution measure of ξ on X {evan: CAN WE SUPPOSE WOLOG THAT $\mu \ll$ nondegenerate Gaussian measures?}.

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

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

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 $\mathcal{H}^n(A) = 0 \implies \mu(P^{-1}_{E,\alpha}(A)) = 0$ for all $A \subseteq E$.

4.1 Geometric Preliminaries

We will be investigating the behavior of $P^{-1}_{E,\alpha}(\cdot)$ 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*

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

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 $(C^\circ)^\circ = \overline{\mathrm{co}(C)}$, where $\mathrm{co}(\cdot)$ denotes the convex hull operation.

Definition 9 (Normal-Regularity[3]). *Let $r \in (0, \infty)$. We say a closed set $E \subseteq X$ is $\underline{r\text{-normal-regular}}$ 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*

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

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 $y \in \mathrm{int}(E)$, 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 {evan: confirm again that this works in infinite-dimensions!}. ☐

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. It seems these definitions do indeed generalize the geometry we are familiar with.

Corollary 1. *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 $\{(1 - \alpha)y + v : \ v \in \widehat{N}_E(y)\}$.

[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.

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

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

$\square$

Corollary 1 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.

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 as we would if μ is a nondegenerate (i.e. positive covariance operator) Gaussian measure on X.

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 $\mathcal{H}^n(S) = 0 \implies P_{E,\alpha}^{-1}(S)$ is Gauss-null. By Corollary 1, we may bound $P_{E,\alpha}^{-1}(S) \subseteq \bigcup_{y \in S}\Big((1-\alpha)y + \widehat{N}_E(y)\Big)$. 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 2 (Uncountable Union of Affine Planes)**.** *Fix $k \in \mathbb{N}$ and $\alpha \in \mathbb{R}$. 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,*

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

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 $X \cong \mathbb{R}^{n+k}$. $\square$

Lemma 3 (Uncountable Union of Affine Half-Planes). *Fix $k \in \mathbb{N}$ and $\alpha \in \mathbb{R}$. 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-dimensional vector half-subspace of X (i.e. up to a unitary, H_x is the set of vectors $(\varphi_1, \varphi_2, \ldots)$ with $\varphi_j = 0$ when $j > k$ and $\varphi_k \geq 0$ in some basis). Then,*

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

Proof. **DO THIS. Boris insists both of these lemmas are true** □

4.3 Main Geometric Result

With the dichotomy given by Lemmas 2 and 3, 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. Do this if we plan to use any of these equivalent formulations. □

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 $k \in \{0, \ldots, n-1\}$.

(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 $k \leq n-1$. Let $\alpha \in \mathbb{R}$ and $S \subseteq E$ be such that $\mathcal{H}^n(S) = 0$. Stratify

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

Define $T_m := S \cap \mathcal{P}_m$ and $\widetilde{T}_k := S \cap \widetilde{\mathcal{P}}_k$ 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 $k \geq n$ 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 1 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 2 completes this direction.

(ii) $\implies$ (i). We will show the contrapositive. Suppose that (i) does not hold, and so there is some $k \in \{0, \ldots, n-1\}$ 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 $k \leq n-1$. At each $y \in \widetilde{\mathcal{P}}_k$, Proposition 3(ii) yields that $T_E(y) \setminus \{0\}$ must be contained in a $(k+1)$-dimensional {evan: **FIX THIS, ITS BROKEN!**} 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 3, 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 $B_r(E) \cap P_{E,0}^{-1}(G)$ has nonempty interior, which means that $P_{E,\alpha}^{-1}(G)$ does too. In the case $\alpha = 1$, then {evan: **prove the $\alpha = 1$ case**} ... 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. $\square$

Remark 4. The condition (i) above 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 $T_E(x)$ 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 Almgren's stratification by dimension of spine of the tangent cone, see Definition 1.6 in these notes). By Theorem 1.7 in the notes, we automatically have that $\dim_{\mathcal{H}}(\widetilde{\mathcal{P}}_k) \leq k$. Thus, for general E we cannot guarantee that $\mathcal{H}^k(\widetilde{\mathcal{P}}_k) = 0$, but we do have that $\mathcal{H}^{k+\delta}(\widetilde{\mathcal{P}}_k) = 0$ for all $\delta > 0$. Furthermore, if we know that E is stationary (w.r.t. $\mathcal{H}^n$ for some $n \geq k$), then $\widetilde{\mathcal{P}}_k$ is k-rectifiable by a result of Naber-Valtorta cited in the aforementioned notes.

5 Returning to Stochastic Processes

Lastly, 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)

Let us 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 4. *Let $f : U \to V$ be a map between subsets of Hilbert spaces. Then,*

$$Df(t) \subseteq T_{f(U)}(f(t)) \subseteq \overline{\mathrm{co}(Df(t))},$$

where the second inclusion also requires U to be compact. Therefore, $f(U)$ is k-pointed at $f(t) \in X$ if and only if $Df(t)$ is k-pointed at 0.

Proof. (First $\subseteq$) 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.

(Second $\subseteq$) Suppose now that $v \in T_{f(U)}(f(t))$ 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

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

{evan: prove this part please :)}

(The rest) The first inclusion tells us the following: if $\overline{\mathrm{co}(T_{f(T)}(f(t)))}$ does not contain any k-dimensional vector subspaces, then neither does $\overline{\mathrm{co}(Df(t))}$. 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 $\overline{\mathrm{co}(T_{f(T)}(f(t)))}$ does contain a k-dimensional subspace, then so does $\overline{\mathrm{co}(Df(t))}$, and so $Df(t)$ is not k-pointed at 0 either. $\square$

Corollary 2. *Let $\Phi : T \to X$ be the spectral feature map for the covariance kernel of a square-integrable, centered stochastic process W with diagonal kernel J_W. Then, $\Phi(T)$ is k-pointed at $\Phi(t)$ if and 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 $u, v \in \mathrm{dom}(D\Phi)(t)$, {evan: **FINISH!** the vibe is $\langle \Phi'(t,u), \Phi'(t,v)\rangle_X = K'_W((t,t);(u,v))$} $\square$

5.2 Wrapping Up

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

Lemma 5. *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**} $\square$

There is one final snag 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 continuity of the Karhunen-Loeve transform.

Proposition 4. *Let μ be the distribution of the random variable ξ from Theorem 2. Then, under Assumption 1, $\nu_1 \ll \mu \ll \nu_2$ for some nondegenerate Gaussian measures ν_1, ν_2 on X.*

Proof. {evan: you know what to do} $\square$

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 condition (i) holds. By Corollary 2, for each $k \in \{1, \dots, n\}$, 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 condition (ii) holds and that $\gamma \ll \mathcal{H}^n \llcorner T$ and J_W is bi-Lipschitz. 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 5, we can apply the converse of Theorem 3 to see that for each $k \in \{1, \dots, n\}$, the set of points $\Phi(t)$ at which $\Phi(T)$ is k-pointed is $\mathcal{H}^{k-1}$-null. (i) follows from Corollary 2. □

7 Conclusion

Loosely speaking, we have shown that the argmax of a broad class of stochastic processes[4] indexed by $\mathbb{R}^n$ is a.c. over the index set if and (with a bit more regularity) 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 and the geometric perspective are of independent interest.

References

[1] link

[2] link

[3] link

[4] this link

[5] da linky

[6] linkster

[4]Processes satisfying Asssumption 1 whose Karhunen-Loeve transforms are equivalent to nondegenerate Gaussian measures on a feature space.