

1 The Ergodicity Metric

In the following discussion, we introduce the concept of the ergodicity metric. Consider a rectangular region $U \subset \mathbb{R}^d$ and assume that the trajectories of the agents are confined within this domain. In [MM21] George Mathew and Igor Mezić defined a dynamical system to be ergodic if, for any Borel-measurable subset of the domain U and any time interval $[0, t]$ the total time spent by a trajectory in that subset is equal to the measure μ of the set. It is noteworthy that the measure μ may be defined in an arbitrary manner, contingent upon the stipulated sampling or exploratory strategy for the domain.. For the sake of intuition we may assume μ to be a probability measure over the domain U . The system is then said to be ergodic if for any set the probability that any agent is within this set at any time, in the given time interval, is equal to the probability measure of this set. Formally, consider a system with a single agent and trajectory $g : [0, t] \rightarrow U$, then we say that the system is ergodic, with respect to the target probability measure μ , if and only if for $\tau \sim \mathcal{U}[0, t]$ and any measurable subset $U' \subset U$, we have that:

$$\mathbb{P}(g(\tau) \in U') = \mu(U') \quad (1.1)$$

holds. For multiple agents with trajectories $g_i, i = 1, 2, \dots, M$ respectively, this readily extends to

$$\frac{1}{M} \sum_{i=1}^M \mathbb{P}(g_i(\tau) \in U') = \mu(U'), U' \subset U \text{ measurable.} \quad (1.2)$$

Now, more formally, we consider the measure spaces $\{U, \Sigma_U, \lambda_g\}$ and $\{[0, t], \Sigma, \lambda\}$, where λ denotes the Lebesgue measure. Let $g : [0, t] \rightarrow U$ be a measurable function (cf. Propostion A.2), then in accordance with Definition A.4 we define the measure induced by the trajectory g on U as $\lambda_g := g(\lambda)$. We now state, that by definition the relation:

$$\lambda_g(U') = \lambda(g^{-1}(U')) = \lambda(\tau \in [0, t] \mid g(\tau) \in U'), U' \in \Sigma_U, \quad (1.3)$$

holds. Intuitively Eq. (1.3) measures the time the trajectory g spends in some set U' . We note that λ_g is not a probability measure as it is t -bounded. However normalizing by the factor $\frac{1}{t}$ makes it such. Before defining the ergodicity metric more formally we need a technical lemma.

Lemma 1.1. *Let $\{\mu_i\}_{i \in \mathbb{N}}$ be a collection of finite measures, where each μ_i is defined on a corresponding σ -algebra \mathcal{A}_i for each $i \in \mathbb{N}$. Then $\mu = \sum_i \mu_i$ is a measure on the σ -algebra $\mathcal{A} = \bigcap_i \mathcal{A}_i$. The measure space (S, \mathcal{A}, μ) enjoys the property that for any measurable function $f : S \rightarrow \mathbb{R}$, the relation:*

$$\int f d\mu = \sum_i \int f d\mu_i \quad (1.4)$$

holds. The symbol $\overline{\mathbb{R}}$ denotes the closure of \mathbb{R} . For convenience, and with some abuse of notation, we write $\overline{\mathbb{R}} = [-\infty, \infty]$. Conventions with regard to the symbol ∞ are discussed in Section B.2.

Proof. See Section B.1 □

We are now ready to define the ergodicity metric.

Definition 1.1. *Let $\{U, \Sigma_U, \mu\}$ be a probability space and $\{[0, t], \Sigma, \lambda\}$ be measure space, then a dynamical system induced by a set of measurable functions $G := \{g_1, g_2, \dots, g_n\}$, $g_j : [0, t] \rightarrow U$, has the ergodic property if and only if*

$$\frac{1}{Nt} \lambda_G = \mu,$$

where λ_G is a measure on U defined under the mappings G , formally $\lambda_G := \sum_{j=1}^N g_j(\lambda) := \sum_{j=1}^N \lambda_{g_j}$ and λ denotes the lebesgue measure.

2 Discussion on Ergodic Solutions

In practical applications, we aim to identify functions G such that the induced dynamical system is ergodic, as defined in Definition 1.1. As a first step, we address the natural question: under what

conditions can such functions be found to ensure ergodicity of the system? To maintain both theoretical clarity and practical relevance, we restrict our analysis to the case where $G := \{g_1, g_2, \dots, g_n\}$ consists of continuous functions. For Borel measures and a single continuous trajectory, the answer is provided by the following theorem.

Theorem 2.1 ([Bog07b, Thm 9.7.1. pg. 289]). *Let K be a compact metric space that is the image of $[0, 1]$ under a continuous mapping f , and let μ be a Borel probability measure on K such that K is its support. Then, there exists a continuous mapping $g : [0, 1] \rightarrow K$ such that $\mu = \lambda \circ g^{-1}$, where λ is Lebesgue measure on $[0, 1]$.*

The following proposition extends the previous Theorem 2.1 to multiple continuous functions G .

Proposition 2.1 (Extension of Theorem 2.1 to multi-agent systems). *Let $\{[0, 1], \mathcal{B}([0, 1]), \lambda\}$ and $\{K, \mathcal{B}(K), \mu\}$ be measure spaces and let $\{K_i\}_{1 \leq i \leq n}$ be compact metric spaces that are the image of $[0, 1]$ under a continuous mapping f_i , such that $\cup K_i = K$ and $\cap K_i = \emptyset$. Now, let μ be a measure on K , such that there exists Borel probability measures μ_i , which are equivalent to μ on K_i , but vanish everywhere else. Then there exists continuous mappings $G := \{g_1, g_2, \dots, g_n\}$, with $g_i : [0, 1] \rightarrow K_i$ such that $\mu = \sum_{i=1}^n \lambda \circ g_i^{-1}$, where λ is the Lebesgue measure on $[0, 1]$. We note that in general $f_i \neq g_i$.*

Proof. By the statements of Theorem 2.1, we have that for $\{[0, 1], \mathcal{B}([0, 1]), \lambda\}$ and $\{K_i, \mathcal{B}(K), \mu\}$, there exists continuous function g such that $\mu_i = \lambda \circ g^{-1}$. Then by Lemma 1.1 and the fact that the K_i 's are disjoint, the assertion is shown. \square

A discussion on the condition of a metric spaces K_i being the image of continuous function on $[0, 1]$ in Proposition 2.1 can be found in Section B.3. We will now give a concrete example of a construction

Example 2.1 (Construction of an trajectory defining an ergodic system). *Let $\{[0, 1], \mathcal{B}([0, 1]), \lambda\}$ and $\{[0, 1]^n, \mathcal{B}([0, 1]^n), \lambda^n\}$ be the measure spaces under consideration, where λ^n is the Lebesgue measure on a n -fold Borel-measurable product space. We note that $[0, 1]^n$ is the image of a continuous function under a continuous mapping f (cf. Theorem B.1), hence a single trajectory $\{g\}$ suffices to achieve ergodicity. An example of such a mapping is the so called Peano curve, of which the graph is illustrated in Figure 1. In fact it can be shown that this function is measure preserving with respect to the Lebesgue measure that is*

$$\lambda \circ f^{-1} = \lambda^n,$$

where $f : [0, 1] \rightarrow [0, 1]^n$ is the appropriate Peano curve, for a proof we refer to [Mil76, Thm 4.62].

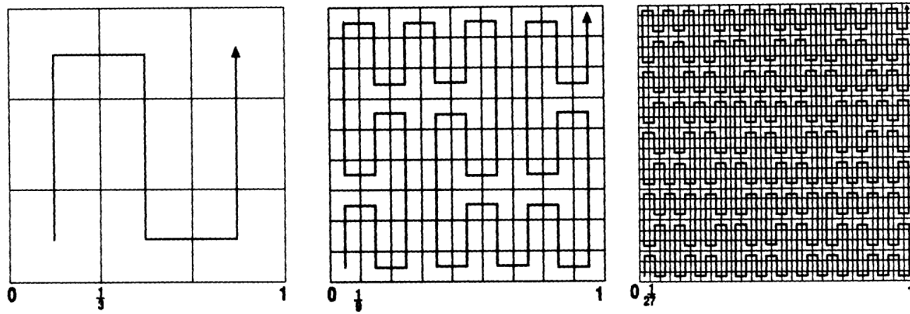


Figure 1: Visualization of the first three elements of a sequence $\{f_k\}_{k \geq 1}$ converging pointwise to the Peano curve. The image was taken from [Sag94, Figure 3.3.2]

Example 2.1 nicely shows that even for the most elementary Borel-measures a system is ergodic only with respect to highly complicated functions, in fact it is well known that every space-filling curve, as the Peano curve has a fractal graph, as well as infinite length. Clearly, no trajectory in finite time can possibly have a infinite length graph. Hence, in general no solution exists. We may then ask ourselves the question if for Borel-measures no suitable solutions exist what about other measures in particular when we restrict the σ -Algebra on the higher dimensional space. As a motivating example consider measure space $\{[0, 1]^n, \{\emptyset, [0, 1]^n\}, \lambda^n\}$, then any trajectory $g : [0, 1] \rightarrow [0, 1]^n$ induces a ergodic system, as in this example a system is ergodic if the trajectory $g : [0, 1] \rightarrow U$ stays within the bounds which

it does by definition, i.e $U \subseteq [0, 1]^n$. Now for measures of the type, as in the previous example the following Proposition guarantees existence of a ****feasible**** solution.

Proposition 2.2. *Consider the measure spaces $\{[0, 1], \mathcal{B}([0, 1]), \lambda\}$ and $\{[0, 1]^n, \mathcal{A}, \mu\}$, where μ is a atom-less measure with support on $[0, 1]^n$ and \mathcal{A} is an Algebra generated by a finite system of closed subsets in $[0, 1]^n$, then there is a continuous function $g : [0, 1] \rightarrow [0, 1]^n$ such that*

$$\mu = \lambda \circ g^{-1}.$$

Proof. Before stating the proof we introduce a technical Lemma.

Lemma 2.1. *Let \mathcal{A} be a Algebra on S generated by a finite cardinality system of closed subsets $\mathcal{E} = \{A_i\}_{i \geq 1}$, then there exist a finite amount of sets $\mathcal{E}' = \{A'_i\}_{i \geq 1}$, such that \mathcal{E} and \mathcal{E}' both generate \mathcal{A} . For $\{A'_i\}_{i \geq 1}$, we require $\bigcup_i A'_i = S$ and $\bigcap_i A'_i = \emptyset$*

Proof. First, we state that if we add elements from an Algebra to a system of subset generating it, the new system still generates said Algebra. Formally, let \mathcal{A} be generated by $\mathcal{E} = \{A_i\}_{i \geq 1}$, then $\{A_i\}_{i \geq 1} \cup \{B_i\}_{i \geq 1}$ generates \mathcal{A} if $\forall_i B_i \in \mathcal{A}$. Hence we can assume without loss of generality that $\bigcup_i A_i = S$. Next define the generator $\mathcal{E}' = \{A'_i\}_{i \geq 1}$, with

$$A'_i = \overline{A_i \setminus \bigcup_{j \neq i} A_j},$$

clearly all $A'_i \in \mathcal{A}$ ¹ and both systems generate the same σ - Algebra. \square

As \mathcal{A} was generated by system of subsets of finite cardinality we can conclude that there exists a partition of closed sets $\{A_i\}_{i \geq 1}$ of $[0, 1]^n$, which generates \mathcal{A} . By a partion of S with closed sets $\{A_i\}_{i \geq 1}$, we mean that $\bigcup_i A_i = S$, $\bigcup_{i: i \neq j} A_i \subset S, j = 1, 2, \dots$ and $\cap \text{int}(A_i) = \emptyset$ (cf. Lemma 2.1). Furthermore, as μ has full support we can assume $\mu(A_i) > 0, i = 1, 2, \dots$. Now by convexity of $[0, 1]^n$ we state that for any $x, y \in [0, 1]^n$, there exist a continuous (linear) function $\gamma : [a, b] \rightarrow [0, 1]^n$, with $\gamma(a) = x$ and $\gamma(b) = y, [a, b] \subseteq [0, 1]$. Now let $\{x_i\}_{i \geq 1}$ be a sequence of points on the boundaries of the A_i 's, such that for each set A_i there exists $x_i, x_{i+1} \in A_i \setminus \text{int}(A_i)$, with a linear path between the points. Such a sequence can be constructed by choosing an arbitrary sequence $\{x_j\}$ satisfying $x_i, x_{i+1} \in A_i \setminus \text{int}(A_i), i = 1, 2, \dots$, then constructing the linear path $\gamma : [0, 1] \rightarrow [0, 1]^n$, with $\gamma(0) = x_i$ and $\gamma(1) = x_{i+1}$, and then joining with the tuple $Y = \{\gamma([0, 1]) \cap A_j \setminus \text{int}(A_j), j \neq i\}$, that is $\{x_1, \dots, x_i, Y, x_{i+1}, \dots\}$. We require that Y is ordered such that $\gamma^{-1}(y_k) < \gamma^{-1}(y_l)$ implies $k < l$. This is well defined as γ is injective. Finally, we define the disjoint partition of $[0, 1] = \bigcup I_j$, such that $\sum_{j: x_j, x_{j+1} \in A_i} \lambda(I_j) = \mu(A_i)$, by construction of the sequence $\{x_i\}_{i \geq 1}$ and the convexity of $[0, 1]^n$, we can conclude existence of a function γ , such that $\gamma^{-1}(A_i) = \bigcup_{j: x_j, x_{j+1} \in A_i} I_j$. The assertion then follows by \cap - stability of $\{A_i\}_{i \geq 1}$ and the statement of Proposition A.5. \square

Remark 2.1. *The length of the graph in Proposition 2.2 is upper bounded by $N \max_i d(x_i, x_{i+1})$, where N denotes the number of disjoint subsets generating the Algebra as per Lemma 2.1.*

3 Defintion of Ergodic Loss functions

In general it is not clear how to compare measures, as simply checking $\lambda_G(A) = \mu(A)$ for all measurable sets is exhaustive. To formulate the approach taken in [MM21] we first need the notion of a measure determining set of linear functionals.

Definition 3.1. *Let $\{S, \mathcal{A}\}$ and $\{S', \mathcal{A}'\}$ be measurable spaces. We say a set of \mathcal{A} - \mathcal{A}' - measurable functions $\{f_k\}_{k \geq 1}$ $f_k : S \rightarrow S'$ is measure determining with respect to the set of measures $M(S)$ if and only if*

$$\forall_k \int f_k d\mu = \int f_k d\lambda \implies \mu = \lambda,$$

where $\mu, \lambda \in M(S)$.

Some examples of measure determining functions are

¹Additional to unions, σ - Algebras are closed under intersections

- For measurable space $\{S, \{S, \emptyset\}\}$ the function $f : S \rightarrow \mathbb{R} : x \mapsto 1$, is measure determining, with respect to the set of finite measures on S , as $\int_S d\mu = \int_S d\lambda \implies \mu(S) = \lambda(S)$. For the empty set $\emptyset = S/S$, we have $\mu(S) - \mu(S) = \lambda(S) - \lambda(S) = 0$
- By extension of the previous example for measurable space $\{S, \{A_1, A_2, \dots, A_n\}\}$, the set of functions $\{f_k\}_{1 \leq k \leq n}$ $f_k = I_{A_k}$ is measure determining.
- For measurable space $\{\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\}$ the set of functions $\{\exp(i\langle t, x \rangle)\}_{t \in \mathbb{R}^n}$ is measure determining, see Proposition 3.1.

Proposition 3.1 ([Bog07a, Proposition 3.8.6, pg. 197]). *If two finite Borel measures have equal Fourier transforms, then they coincide. In particular, two integrable functions with equal Fourier transforms are equal almost everywhere.*

We now state a necessary and sufficient condition for a set of functions being measure determining over all finite measures on a compact set.

Proposition 3.2. *Let $M(X)$ denote the set of all finite Borel-measures (possibly complex) over X . Then $\{f_k\}$ is measure determining if and only if $C := \text{span}(\{f_k\}_k)$ is dense in $L^1(\mu, X)$, $\mu \in M(X)$.*

Proof. For sufficiency we consider $\lambda = \mu - \nu$, we then need to show that

$$\forall f \in C \int f d\lambda = \int_X f d\mu - \int_X f d\nu = 0 \implies \lambda = 0, \quad (3.5)$$

by the finiteness of the respective measures this integral always exists and λ is finite aswell. Now as C is dense in L^1 there exists a sequence $\{f_n\}_{n \geq 1}$ such that $f_1 \leq f_2 \leq \dots$ and $\sup_{n \geq 1} I_A$ for any measurable set A . Now by Proposition A.6 we can write

$$\int I_A d\lambda = \lim_{n \rightarrow \infty} \int f_n d\lambda = \lim_{n \rightarrow \infty} \left(\int f_n d\mu - \int f_n d\nu \right) = \int I_A d\mu - \int I_A d\nu = 0,$$

hence μ and ν agree on all measurable sets, showing the implication.

Assume for the sake of contradiction that the set of functions $\{f_k\}$ is measure-determining, but its linear span C , is not dense in L^1 . Then, by the Hahn–Banach theorem [Rud91, Thm. 3.5], there exists a non-zero continuous linear functional

$$T \in (L^1(X))^*$$

such that

$$T(f) = 0 \quad \text{for all } f \in \overline{C},$$

where \overline{C} denotes the closure of C . By the Riesz Representation Theorem [Rud87, Thm 6.19], every continuous linear functional on $L^1(X)$ can be represented by a signed (or finite) Borel measure μ on X , so that

$$T(f) = \int_X f d\mu \quad \text{for all } f \in L^1(X).$$

Hence,

$$\int f d\mu = 0 \quad \text{for all } f \in \overline{C},$$

but since $T \neq 0$, we have $\mu \neq 0$. However, μ agrees for all $f \in \overline{C}$ with the zero measure. Contradiction! \square

For the previous definition we may readily deduce a more convenient form to measure the ergodicity of a dynamical system, which is given in the following proposition.

Proposition 3.3. *Let $\{f_k\}_{k \geq 1}$ be a set of measure determining functions as per Definition 3.1, then the dynamical system induced by a set of measurable functions $G := \{g_1, g_2, \dots, g_n\}$, $g_j : [0, t] \rightarrow U$ has the ergodic property if and only if*

$$\forall_k \int f_k d\lambda_G = \sum_{j=1}^n \int f_k \circ g_j d\lambda \quad (3.6)$$

holds. Here λ denotes the Lebesgue measure and λ_G is a measure on U defined under the mappings G , formally $\lambda_G := \sum_{j=1}^N g_j(\lambda) := \sum_{j=1}^N \lambda_{g_j}$.

Proof. Before beginning the proof we state a technical Lemma.

Lemma 3.1 ([Bog07a, Thm 3.6.1]). *Let $(S, A, \lambda), (S', A', \lambda')$ be measure spaces and λ denote the Lebesgue measure. Let g be measurable and define $\lambda' = g(\lambda)$ (cf. Definition A.4 and Eq. (1.3)). Then a function f is integrable with respect to λ' if and only if $\int f \circ g d\lambda < \infty$, i.e the composition of f and g is integrable. In which case*

$$\int f d\lambda' = \int f \circ g d\lambda$$

holds.

Then by an simple application of the Lemmata 1.1 and 3.1 to left hand side of Eq. (3.6) we obtain

$$\begin{aligned} \int f_k d\lambda_G &= \sum_{j=1}^n \int f_k d\lambda_{g_j} \text{ by Lemma 1.1} \\ &= \sum_{j=1}^N \int f_k \circ g_j d\lambda \text{ by Lemma 3.1.} \end{aligned} \tag{3.7}$$

Finally, the assertion follows by the measure determining property of functions $\{f_k\}$. \square

For a discussion of the connection between this framework and [MM21] we refer the reader to Section B.4. Finally, given that, as stated in Section 2, we are not able to define a feasible ergodic solution given an underlying Borel- σ algebra, we are now able to define a metric that provides an estimate of how far a given trajectory is from being ergodic, that is $\sum_k l(\langle f_k, \lambda \rangle, \langle \lambda_G, f_k \rangle)$, where λ_G describes the measure on U defined under the mappings G , λ denotes the target measure and l denotes an arbitrary loss function. We note that structurally different types of loss functions have already been employed successfully. For example, in [SGTM25], the authors use the inner product loss $\langle \lambda - \lambda_G, \lambda - \lambda_G \rangle$ with great success.

4 Optimization

We now restate the key results from [MM21] concerning the optimization of the ergodic metric expressed in its measure-determining functional form.

Consider the cost function defined as

$$\Phi(t) = \frac{1}{2} \sum_k (\langle f_k, \lambda_G \rangle - |G| \langle f_k, \lambda \rangle)^2 = \frac{1}{2} \sum_k |S_k(t)|^2,$$

where $S_k(t) := \langle f_k, \lambda_G \rangle - |G| \langle f_k, \lambda \rangle$.

To facilitate optimization, we perform a second-order Taylor expansion of Φ around time t , yielding

$$\Phi(t + \Delta t) = \Phi(t) + \dot{\Phi}(t) \Delta t + \frac{1}{2} \ddot{\Phi}(t) (\Delta t)^2 + \mathcal{O}((\Delta t)^3).$$

The first and second derivatives of Φ can be expressed as

$$\dot{\Phi}(t) = \sum_k S_k(t) \dot{S}_k(t),$$

and

$$\ddot{\Phi}(t) = \sum_k \dot{S}_k(t)^2 + \sum_k S_k(t) \ddot{S}_k(t).$$

The derivatives of S_k are given by

$$\dot{S}_k = \sum_{g \in G} \langle f_k, \lambda_{g(\tau)} \rangle - |G| \langle f_k, \lambda \rangle,$$

and

$$\ddot{S}_k = \sum_{g \in G} \nabla \langle f_k, \lambda_{g(\tau)} \rangle Dg,$$

where $Dg = f(g, u)$ represents the state derivative of the agents g , which depends on the control input u .

Substituting these expressions back into the expansion, we obtain

$$\ddot{\Phi}(t) = \sum_k \dot{S}_k(t)^2 + \sum_k S_k(t) \sum_{g \in G} \langle \nabla \langle f_k, \lambda_g \rangle, f(g, u) \rangle.$$

Since only the second term depends explicitly on the control u , the optimization problem reduces to minimizing

$$\sum_k S_k(t) \sum_{g \in G} \langle \nabla \langle f_k, \lambda_{g(\tau)} \rangle, f(g, u) \rangle,$$

subject to the constraint

$$\|u_g\| \leq u_{\max}.$$

In the case of first-order dynamics, where $Dg = u$, this expression simplifies, and the optimal control can be found by solving

$$\min_u \sum_k S_k(t) \sum_{g \in G} \langle \nabla \langle f_k, \lambda_g \rangle, u \rangle, \quad \text{with} \quad \|u\| \leq u_{\max}.$$

This linear optimization is minimized by selecting the control

$$u_g^* = -u_{\max} \frac{\sum_k \nabla \langle f_k, \lambda_g \rangle}{\|\sum_k \nabla \langle f_k, \lambda_g \rangle\|}.$$

We note that even higher order dynamics can be optimized in this fashion but more effort is required for deriving feedback laws.

5 Examples of measure determining sets of functions $\{f_k\}$

In conjunction with Proposition 2.2 the set of indicator function $\{I_{A_k}\}_{1 \leq k \leq N}$, is measure determining over the Algebra $\{A_1, A_2, \dots, A_N\}$. In accordance with the results derived in Section 4 it is evident that these basis are unfeasible for the task at hand, as their gradient evaluates to zero almost everywhere. However, we may still approximate them by a smooth function. That is let $\{X, d_X\}$ be a metric space and $A \subset X$ be a set on which we wish to approximate the Indicator function we then define $A_\epsilon = \{x \in X \mid \text{dist}(x, A) \leq \epsilon\}$, where $\text{dist}(x, A) = \sup_{y \in A} d_X(x, y)$. We write for the approximation of the indicator function over A_ϵ :

$$\tilde{I}_{A_\epsilon}(x) = \begin{cases} 1 & \text{if } x \in A \\ h(x) & \text{if } x \in A_\epsilon \setminus A, \\ 0 & \text{else} \end{cases} \quad (5.8)$$

where h is chosen such that \tilde{I} is smooth. As an example for $A = [a, b] \subset \{\mathbb{R}, |\cdot|\}$ we get

$$\tilde{I}_{[a-\epsilon, b+\epsilon]}(x) = \begin{cases} 1 & \text{if } x \in [a, b] \\ g(x) & \text{if } x \in [b, b+\epsilon] \\ h(x) & \text{if } x \in [a-\epsilon, a] \\ 0 & \text{else.} \end{cases}$$

Intuitively this definition allows for a smooth transition between the respective sets as the boundaries are smooth. An example of such an approximation can be observed in Figure 2.

Smooth Indicator Functions for 5 Adjacent intervals of length 0.2 $\varepsilon = 0.1$

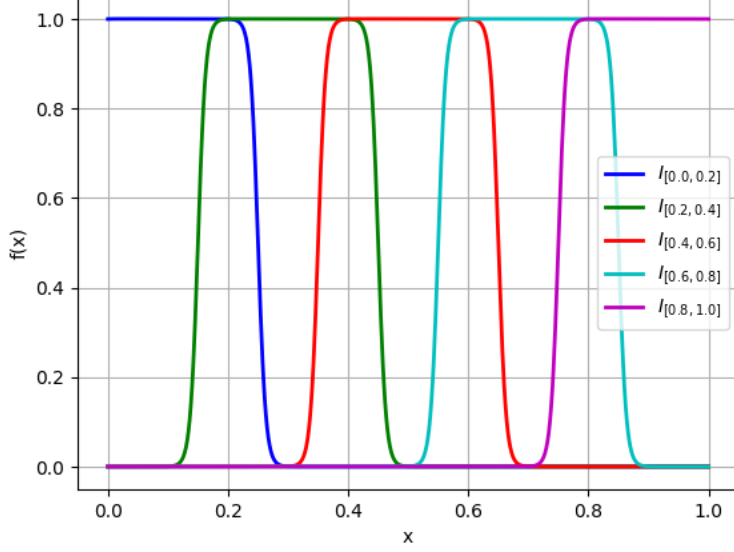


Figure 2: Smooth estimation of indicator functions

As a matter of fact one may understand the smooth approximation of the indicator functions as a linear combination of elementary functions converging uniformly. The smooth approximation of the indicator function as in Eq. (5.8) can be written as

$$f_{N,\varepsilon} = \sum_{i=0}^N c_i I_{A_{\frac{\varepsilon i}{N}}}$$

where $\{c_i\}_{i \in \mathbb{N}}$ is a decreasing sequence, such that $\lim_{i \rightarrow \infty} c_i = 0$ and $\sum_i c_i = 1$ and $A_0 = A$. We then have

$$\lim_{N \rightarrow \infty} f_{\varepsilon,N} = \tilde{I}_{A_\varepsilon},$$

where the $f_{\varepsilon,N}$ converges point-wise to a smooth function, with respect to N . Additionally, we have

$$\lim_{\varepsilon \rightarrow 0} f_{N,\varepsilon} = I_A.$$

the sequence c_i can be understood as a notion of importance of matching at the respective subset of A_ε the greatest importance is given to the set A . A visualization of this can be seen in Figure 3.

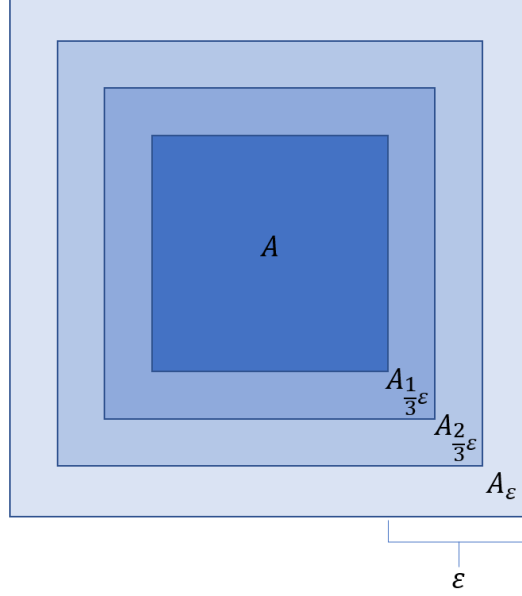


Figure 3: Representation of the sets on which $f_{4,\epsilon}$ takes a unique value. The shade of each set represents the magnitude c_i of the function over this set the brighter the shade the lower the value. The sets are contained within the normed linear space $\{\mathbb{R}^2, \|\cdot\|_\infty\}$

A natural question that arises is whether we can choose sets A and ϵ in such a way that they yield the measure-determining property over an arbitrary algebra. In the case of the Borel- σ -algebra, we will derive the indicator function of a single point from a family of sets. Specifically, we consider sets centered at a collection of points $\{c_i\}$, defined by

$$A_{\epsilon, c_k} = \{x \in X \mid \|x - c_k\|_p \leq \epsilon\}.$$

Figure 3 illustrates such sets using the L_∞ norm. We now proceed to analyze how these sets can be used to construct the indicator function of a point, for that matter consider:

$$f_{c_k, N, \epsilon} = \sum_{i=0}^N c_i I_{A_{\frac{\epsilon_i}{N}}},$$

then there exists some sequence $c_i, i \in \mathbb{N}$, such that $f_{c_k, N, \epsilon}$ converges point wise to a function of the form $f_{c_k} : \mathbb{R}^n \rightarrow \mathbb{R} : x \mapsto \varphi(\|x - c_k\|_2^2)$, where $\varphi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is strictly decreasing. The set of functions $\{f_{c_k}\}_{k \in \mathbb{N}}$ are called the radial basis function. In fact it can be shown that these basis form a dense space in the continuous functions over a compact domain for $p \in [1, 2]$ [Bax10]. We can therefore conclude, by Proposition 3.2 together with the fact that the continuous functions form a dense subspace of L_1 , that $\{f_{c_k}\}$ is measure determining on the Borel σ -Algebras. In conclusion, we have motivated the use of radial basis functions as approximations of indicator functions centered at single points.

Specifically, we propose the following sets of functions to approximate the measure-determining property—emphasizing that this is only an approximation, as computing an infinite number of coefficients is generally infeasible. These sets include smooth approximations of indicator functions over certain regions, as previously motivated by Lemma 2.1. In particular, when the underlying sets are spherical or rectangular, this yields a family of measure determining radial basis functions that we also propose as an interesting set of functions. Finally, supported by [MM21] and Proposition 3.1, we include the Fourier basis functions as another set of measure-determining functions.

6 Simulation

In this section we evaluate the efficacy of the introduced measure determining sets of basis functions (cf. Section 5) using the methods introduced in Section 4. For the evaluation we shall run the optimization

as per Section 4 and then project the computed coefficients $c_k = \langle f_k, \lambda_G \rangle$ onto the set of basis functions \tilde{c}_k . Finally we shall evaluate the resulting measure defined by $\lambda_G(A) = \langle I_A, \sum_k \tilde{c}_k f_k \rangle$. We then use the following notions as measure of performance [Ope25].

- **L^2 distance** between two probability density functions (pdfs) p and q :

$$L^2(p, q) = \|p - q\|_2 = \sqrt{\int (p(x) - q(x))^2 dx}$$

Measures the root mean squared difference between the densities. It is sensitive to large point-wise differences but treats all discrepancies uniformly. However, it may not capture perceptual similarity well if differences occur in low-density regions.

- **Total Variation (TV) distance:**

$$\text{TV}(p, q) = \frac{1}{2} \int |p(x) - q(x)| dx$$

Represents the maximal difference in probability mass assigned to any event. It has a clear probabilistic interpretation and is bounded between 0 and 1, but can be overly sensitive to small local fluctuations.

- **Kullback-Leibler (KL) divergence** from p to q :

$$\text{KL}(p||q) = \int p(x) \log \frac{p(x)}{q(x)} dx$$

Measures the information lost when q is used to approximate p . It is asymmetric and heavily penalizes regions where q underestimates p . Useful for inference and likelihood-based methods but undefined if $q = 0$ where $p > 0$.

- **Jensen-Shannon (JS) divergence** (a symmetric and smoothed variant of KL):

$$m(x) = \frac{p(x) + q(x)}{2}, \quad \text{JS}(p, q) = \frac{1}{2} \text{KL}(p||m) + \frac{1}{2} \text{KL}(q||m)$$

Provides a symmetric and finite divergence measure, smoothing the KL divergences by averaging. It is more robust to zero probabilities and interpretable as a measure of similarity.

- **Cosine similarity:**

$$\text{CosSim}(p, q) = \frac{\int p(x)q(x) dx}{\sqrt{\int p(x)^2 dx} \sqrt{\int q(x)^2 dx}}$$

Measures the angle between p and q when viewed as vectors in function space, capturing similarity in shape rather than magnitude. It ranges from 0 (orthogonal) to 1 (identical up to scaling) and is insensitive to scale differences.

Figure 4 illustrates the performance of the different sets of basis functions with respect to the previously introduced evaluation metrics. It is evident that both the radial basis functions and the smooth indicator functions achieve faster convergence compared to the Fourier basis. This behavior is expected, as the Fourier basis functions are not equally weighted across the domain.

Notably, although the Fourier basis functions exhibit slower initial progress, they continue to improve with extended optimization (beyond what is shown in Figure 4), whereas the radial basis and smooth indicator functions tend to plateau quicker. Therefore, we conclude that while the Fourier basis ultimately yields the best results given sufficient time, the radial and indicator-based approaches offer superior performance over shorter time horizons. This property is particularly advantageous in time-critical applications such as search and rescue missions involving moving targets or adaptive target measures, where rapid convergence is essential.

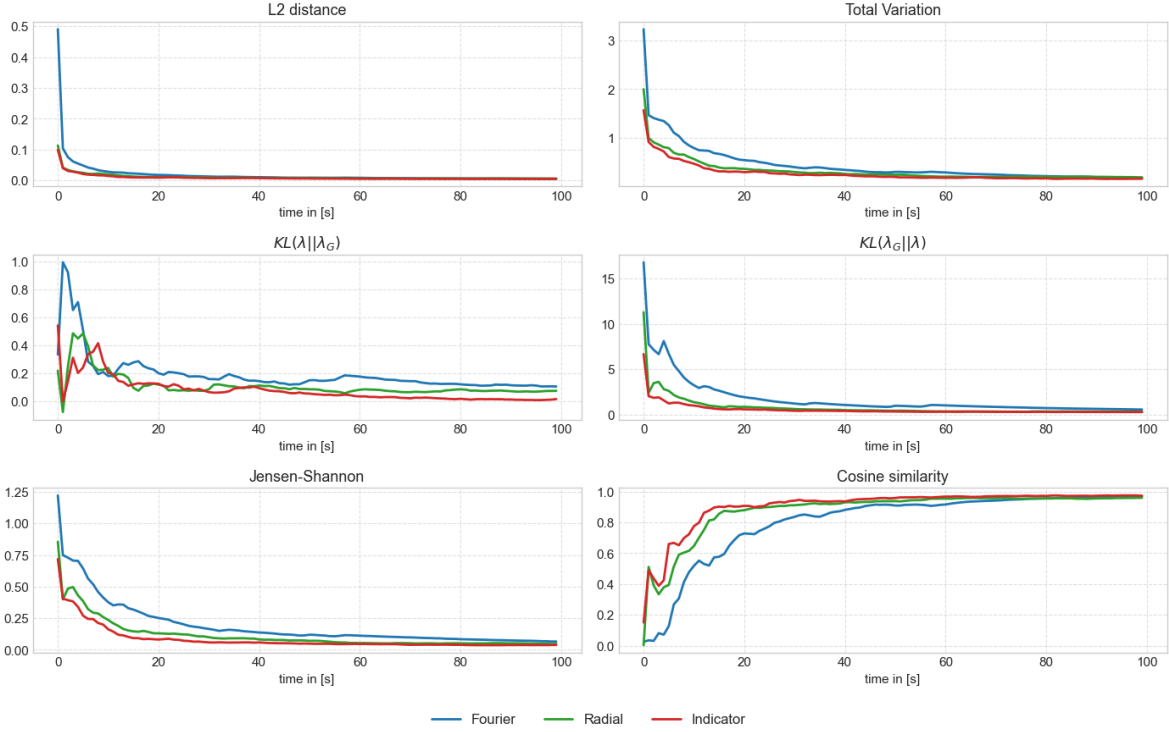


Figure 4: Several evaluation metrics for comparing similarity between probability measures over time on a $[0, 1] \times [0, 1]$. Each set of basis functions consists of 50×50 individual basis. The target measure is described by a gaussian mixture model. The system is governed by first order dynamics with $u_{max} = 0.2$

Another intriguing property of the derived alternative basis functions is that they extend in a more straight forward fashion to non euclidean space, such as $SE(2)$, $SO(3)$ or in general $SU(n)$, which are commonly used in control theory. In [MM13] the extension of the problem to $SE(2)$ using Fourier-basis has been conducted in [SGTM25] however it is stated that in general this extension is highly non trivial using Fourier-basis. Giving superiority to the other types of basis introduced in this work.

We end this Section by noting that proof of Proposition 2.2 is in fact constructive, we leave deriving suitable feedback laws, for example by constructing the paths between the boundaries of the respective sets by Dubin paths, as an probably intriguing problem for future research.

7 Generalization of the ergodic framework

In the following we shall give a more general definition of the ergodicity-metric. To this end, we extend the ergodic measure to incorporate sensor information. Formally, we arrive at the following definition.

Definition 7.1 (Measure-valued Trajectories). *Let $\{U, \mathcal{B}(U)\}$ be a measurable space and let $\mu : U \rightarrow \mathcal{P}(U)$ ² : $x \mapsto \mu_x$ be a weak*-measurable (cf. Definition A.1), measure valued function. Additionally, let $g : [0, t] \rightarrow U$, be the continuous trajectory of an agent. The agent is then said to induce the measure $t \mapsto \frac{1}{t} \int_{[0,t]} \mu_{g(\tau)} d\tau := \lambda_g$ at time t .*

If the weak*-measurability assumption is neglected, that is we take measurability with respect to the strong topology we would exclude important examples, as outlined in the following remark.

Remark 7.1. *The map $\mu : U \rightarrow \mathcal{P}(U)$ needs to be weak* measurable (cf. Definition A.1). As for example the measure valued mapping $t \mapsto \delta_{g(t)} := \delta_x \circ g(t)$ is discontinuous at every point t where g*

² $\mathcal{P}(U)$ denotes the set of Borel probability measures over U

is not locally constant. If the set of such discontinuities has nonzero measure, the mapping fails to be strongly measurable. In contrast, under the weak-* topology, the mapping $t \mapsto \delta_{g(t)}$ is measurable whenever $t \mapsto g(t)$ is measurable.

Proposition 7.1. *The measure μ , induced by an agent, as per Definition 7.1, indeed defines a Borel-probability measure.*

Proof. Monotonicity and σ -continuity are immediate by the fact that each μ_x is a valid measure. The σ -sub additivity follows for disjoint $\{A_i\}$, per the relation

$$\begin{aligned} \mu\left(\bigcup_i A_i\right) &\leq \sum_i \mu(A_i) \\ \frac{1}{t} \int_{[0,t]} \mu_{g(\tau)}\left(\bigcup_i A_i\right) d\tau &\leq \sum_i \frac{1}{t} \int_{[0,t]} \mu_{g(\tau)}(A_i) d\tau \\ \int_{[0,t]} \mu_{g(\tau)}\left(\bigcup_i A_i\right) d\tau &\leq \int_{[0,t]} \sum_i \mu_{g(\tau)}(A_i) d\tau, \end{aligned}$$

which yields the assertion since each $\mu_{g(\tau)}$ is a valid measure. The change of summation and integration on the right hand side is justified by Tonellis theorem A.1. We can employ the theorem as $\tau \mapsto \mu_{g(\tau)}(A)$ is non-negative, bounded and measurable on the finite interval $[0, t]$. This follows from the fact that

$$\tau \mapsto \mu_{g(\tau)}(A),$$

is measurable as per the weak*-measurability assumption and

$$x \mapsto \int_A 1 d\mu_x = \mu_x(A),$$

is measurable. Finally, $g : [0, t] \rightarrow U$ is measurable by continuity and the composition of two measurable functions is measurable. \square

Proposition 7.2. *Let λ_g be a measure; induced by an agent, as per Definition 7.1, then we have that for any measurable function $f : U \rightarrow \mathbb{R}$ with $\int_0^t \int_U |f(x)| d\mu_{g(\tau)}(x) d\tau < \infty$, the following holds:*

$$\int_U f(x) d\lambda_g(x) = \int_0^t \int_U f(x) d\mu_{g(\tau)}(x) d\tau.$$

Proof. 1. *Simple functions.* Let $s(x) = \sum_{k=1}^N a_k \mathbf{1}_{A_k}(x)$ be a simple function. Then

$$\int_U s d\lambda_g = \sum_{k=1}^N a_k \lambda_g(A_k) = \sum_{k=1}^N a_k \int_0^t d\mu_{g(\tau)}(A_k) d\tau = \int_0^t \sum_{k=1}^N a_k d\mu_{g(\tau)}(A_k) d\tau = \int_0^t \int_U s d\mu_{g(\tau)} d\tau,$$

change of integration and summation follows exactly as in the proof of Proposition 7.1.

2. *Nonnegative measurable f .* Given $f \geq 0$, choose an increasing sequence of simple functions $s_n \uparrow f$. By monotone convergence,

$$\int_U f d\lambda_g = \lim_{n \rightarrow \infty} \int_U s_n d\lambda_g = \lim_{n \rightarrow \infty} \int_0^t \int_U s_n d\mu_{g(\tau)} d\tau = \int_0^t \int_U f d\mu_{g(\tau)} d\tau,$$

exactly as in the proof of Lemma 1.1.

3. *General integrable f .* If f satisfies $\int_0^t \int_U |f| d\mu_{g(\tau)} d\tau < \infty$, write $f = f^+ - f^-$ and apply the above to f^+ and f^- . Subtracting gives the desired identity. \square

With these new definitions in mind we aim to explore an alternative approach to trajectory optimization in ergodic exploration by elaborating on the unicycle example presented in [MM13]. In particular, we consider a kinematic model frequently used to represent mobile robotic agents with non holonomic constraints. The state of the system is defined as

$$g(t) = \begin{bmatrix} X(t) \\ Y(t) \\ \theta(t) \end{bmatrix},$$

where $X(t)$ and $Y(t)$ represent the Cartesian coordinates of the agent in the plane, and $\theta(t)$ denotes the agent's heading angle measured counterclockwise from the global X -axis.

The control input is given by

$$u(t) = \begin{bmatrix} v(t) \\ \omega(t) \end{bmatrix},$$

where $v(t)$ is the translational (forward) velocity and $\omega(t)$ is the angular (rotational) velocity.

The time evolution of the state is governed by the unicycle model:

$$\dot{g}(t) = \begin{bmatrix} \cos(\theta(t)) & 0 \\ \sin(\theta(t)) & 0 \\ 0 & 1 \end{bmatrix} \cdot u(t) := M(t)u(t).$$

Additionally we define the measure valued function the agent induces at position x (cf. Definition 7.1) as:

$$\mu_x : (x, y, \theta) \mapsto \frac{1}{\pi r^2} I_{B(\{x+r \cos(\theta), y+r \sin(\theta)\}, r)} \lambda, \quad (7.9)$$

where λ denotes the Lebesgue measure and $B(x, r) = \{y \in U : \|x - y\|_2 \leq r\}$. Then following the steps outlined in Section 4 we obtain

$$\min_u \sum_k S_k(t) \sum_{g \in G} \nabla \langle B_g(t), Dg \rangle, \quad \text{with} \quad \|u\| \leq u_{\max},$$

where $B_g(t) = \langle f_k, \lambda_g \rangle$. This linear optimization is minimized by selecting the control

$$u_g^* = -u_{\max} \frac{\sum_k \langle \nabla B_g(t), M(t) \rangle}{\|\sum_k \langle \nabla B_g(t), M(t) \rangle\|}.$$

Figure 5 shows the results obtained by this feedback law, where the set $\{f_k\}$ was choose to consist of Fourier-basis as in [MM21]. Readers who are concerned with the analytical form of $\langle f_k, \lambda_g \rangle$ are referred to [GS64, pg. 198]

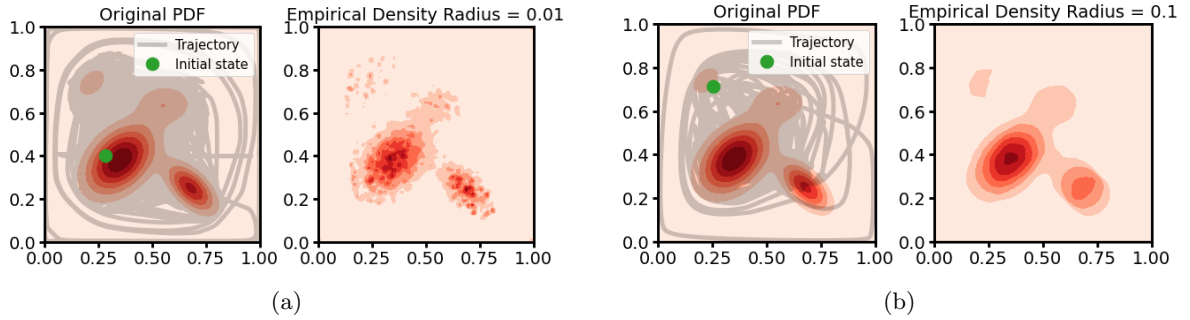


Figure 5: Optimization over 100000 steps. The empirical distribution was obtained by employing Monte Carlo integration to the measure defined in Definition 7.1

With the previously introduced formulation (cf. Definition 7.1), a key challenge arises: the inner product $\langle f_k, \lambda_g \rangle$ may not be available in analytical form for a general measure-valued function μ_x , unlike the special case $\mu_x = \delta_x$.

In scenarios where, for any x , the support of μ_x is small compared to the support of the target measure ν , i.e., $\text{supp}(\mu_x) \ll \text{supp}(\nu)$, a uniform, geometrically motivated approximation of the underlying density of μ_x appears reasonable. However, when the support of the sensor measure μ_x is significant relative to that of the target, this approximation becomes unreliable. Moreover, numerical computation of such integrals would be computationally expensive and would render the application of the ∇ -operator impractical.

Assuming that the mapping μ_x assigns structurally equivalent measures to each point x — for instance, as in Eq. (7.9) — one can exploit standard properties of the Fourier transform. We now introduce the relevant properties of the Fourier transform that enable this approach.

- **Translation (Shift) Theorem:** Translating a function $f(x, y)$ in space by a vector (x_0, y_0) results in a phase shift in the Fourier domain:

$$f(x - x_0, y - y_0) \xleftrightarrow{\mathcal{F}} \hat{f}(k_x, k_y) \cdot e^{-i2\pi(k_x x_0 + k_y y_0)}.$$

This means that once the Fourier coefficients of a function centered at the origin are computed, translating the function simply requires multiplying the spectrum by a complex exponential phase factor.

- **Rotation Theorem:** Rotating a function $f(x, y)$ by an angle α corresponds to rotating its Fourier transform by the same angle:

$$f(R_\alpha^{-1} \mathbf{x}) \xleftrightarrow{\mathcal{F}} \hat{f}(R_\alpha^{-1} \mathbf{k}),$$

where the rotation matrix is given by

$$R_\alpha = \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix}.$$

In practice, this means that if a function is rotated in the spatial domain, its frequency content is also rotated accordingly in the Fourier domain.

Together, these properties imply that once the Fourier transform of a base function (e.g., centered and unrotated) is computed, the transform of any translated and rotated version can be obtained analytically by combining phase modulation and frequency-space rotation. The Fourier coefficients of the underlying density rotated by an angle α and translated to the point (x_0, y_0) are then given by the following expression:

$$\hat{f}_{transformed}(\mathbf{k}) = \hat{f}(R_\alpha^{-1} \mathbf{k}) e^{-i2\pi(k_x x_0 + k_y y_0)} \quad (7.10)$$

This avoids recomputing the full transform and enables efficient gradient computations with respect to the translation and rotation parameters. The gradient with respect to $\{x_0, y_0, \alpha\}$ is then given by:

$$\nabla \hat{f}_{transformed}(k_x, k_y) = e^{-i2\pi(k_x x_0 + k_y y_0)} \cdot \begin{bmatrix} -i2\pi k_x \cdot \hat{f}(R_\alpha^{-1} \mathbf{k}) \\ -i2\pi k_y \cdot \hat{f}(R_\alpha^{-1} \mathbf{k}) \\ \nabla_{\mathbf{k}'} \hat{f}(\mathbf{k}') \cdot \frac{\partial \mathbf{k}'}{\partial \alpha} \end{bmatrix},$$

$$\text{where } \mathbf{k} = \begin{bmatrix} k_x \\ k_y \end{bmatrix}, \quad \mathbf{k}' = R_{-\alpha} \mathbf{k}, \quad R_{-\alpha} = \begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix}, \quad \frac{\partial \mathbf{k}'}{\partial \alpha} = \begin{bmatrix} -k_x \sin \alpha + k_y \cos \alpha \\ -k_x \cos \alpha - k_y \sin \alpha \end{bmatrix}$$

Accordingly, the results of Section 4 can be utilized. It is imperative to recognize that, in most cases, the function $f(\mathbf{k})$ is not available and must be estimated. Before leveraging the method introduced in Section 4, we shall also take into consideration the imaginary parts of the respective Fourier coefficients by incorporating them into the loss function. To this end, we define the new loss function as:

$$\Phi(t) = \frac{1}{2} \sum_k (\Re[\langle f_k, \lambda_g \rangle] - \Re[t \langle f_k, \lambda \rangle])^2 + \frac{1}{2} \sum_k (\Im[\langle f_k, \lambda_g \rangle] - \Im[t \langle f_k, \lambda \rangle])^2,$$

where $f_k = e^{-i2\pi(k_x x + k_y y)}$ and λ denotes the target measure. At this juncture, it is possible to leverage the results from Section 4 directly to derive the feedback law:

$$u_g^* = -u_{\max} \frac{\sum_k \nabla \Re[\langle f_k, \lambda_g \rangle] + \nabla \Im[\langle f_k, \lambda_g \rangle]}{\|\sum_k \nabla \Re[\langle f_k, \lambda_g \rangle] + \nabla \Im[\langle f_k, \lambda_g \rangle]\|}.$$

The following measure-valued function, characterized by an angle-dependent and non-symmetric density, will be used to evaluate the efficiency of the proposed approach:

$$(x, y, \theta) \mapsto f(x, y, \mu_\theta) \lambda, \quad (7.11)$$

where $f(x, y, \mu_\theta) = \frac{1}{2\pi\sigma^2 I_0(\kappa)} \exp\left(-\frac{(x-x_0)^2 + (y-y_0)^2}{2\sigma^2}\right) \cdot \exp(\kappa \cos(\theta(x, y) - \mu_\theta))$,
with $\theta(x, y) = \arctan 2(y - y_0, x - x_0)$

and I_0 denotes the modified Bessel-function of the first kind. The rotation and translation of this function to an arbitrary point in the domain is visualized in Figure 6.

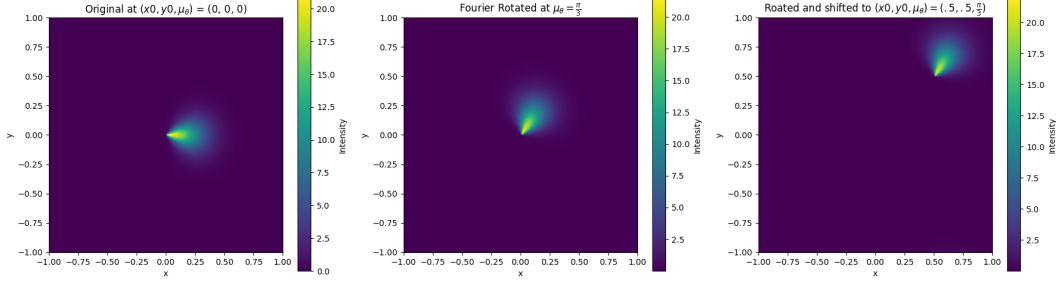


Figure 6: Rotation and translation of the objective function in the Fourier domain

Figure 7 illustrates the trajectory generated by the proposed method. This trajectory demonstrates how the system evolves over time under the derived feedback law, highlighting the effectiveness of our approach in guiding the dynamics towards the target distribution.

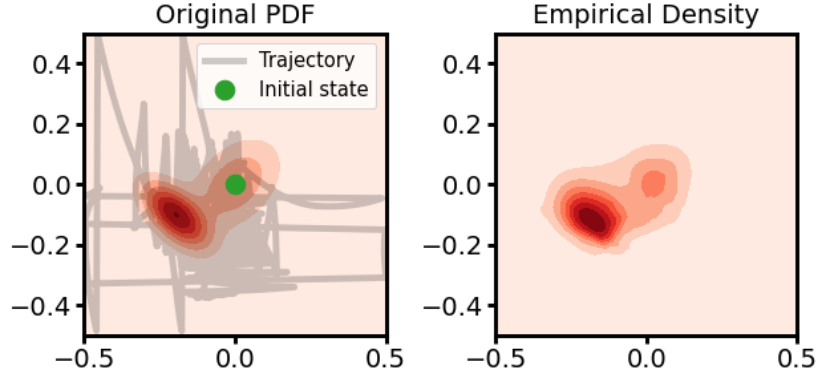


Figure 7: Sampling a Gaussian mixture density using the sensor measure defined in Eq. 7.11 with parameters $\kappa = 5$ and $\sigma = 0.1$. The empirical density was obtained via Monte Carlo integration of the Fourier coefficients followed by the inverse transform. The trajectory was simulated over 50,000 timesteps, each corresponding to 0.01 seconds.

8 Conclusion

In this work, we rigorously introduced the concept of measure-determining sets of functions. We also established a necessary and sufficient condition for these sets to determine finite Borel measures. In this context, we examined three notable classes of functions: Fourier basis functions, radial basis functions, and smooth approximations of indicator functions. To evaluate their practical applications, we analyzed how the ergodic metric converges when using these sets. Our findings suggest that Fourier basis functions exhibit superior asymptotic convergence characteristics, whereas radial and smoothed indicator functions converge more quickly initially but saturate sooner. This behavior suggests their particular suitability for time-sensitive applications, such as the pursuit of dynamically evolving target measures, where rapid exploratory behavior is prioritized over long-term thoroughness. In the second part of this paper, we extended the previously defined ergodic metric to incorporate sensor-based information in a mathematically rigorous manner. Building on this extension, we generalized the associated feedback control laws to accommodate sensor-dependent formulations. To address the challenges posed by highly nontrivial sensor measurements, we proposed a principled, computationally

efficient solution. We demonstrated the effectiveness of this approach through a representative example, highlighting its practical applicability in complex sensing environments.

A Auxiliary results

A.1 Measurability

Proposition A.1 (Generated σ -algebras [BK13, pg. 8]). *For every system \mathcal{E} of subsets in S , there exists a smallest σ -algebra which contains \mathcal{E} . We obtain it as the intersection of all σ -algebras containing \mathcal{E} :*

$$\sigma(\mathcal{E}) = \{A \subset S \mid A \subset \tilde{A} \text{ for every } \sigma\text{-algebra } \tilde{A} \text{ in } S \text{ with } \tilde{A} \supset \mathcal{E}\}$$

Proposition A.2 (Measurability criterion [BK13, pg. 9]). *Let (S, \mathcal{A}) and (S', \mathcal{A}') be measurable spaces, and let \mathcal{E}' generate \mathcal{A}' . Then a mapping $\varphi : S \rightarrow S'$ is \mathcal{A} - \mathcal{A}' -measurable if*

$$\varphi^{-1}(A') \in \mathcal{A} \quad \text{for all } A' \in \mathcal{E}'.$$

Definition A.1 (Weak Measurability). *Let (Ω, \mathcal{F}) be a measurable space and E a Banach space. A function $f : \Omega \rightarrow E$ is said to be weakly measurable if for every continuous linear functional $\varphi \in E^*$, the scalar function*

$$\omega \mapsto \langle \varphi, f(\omega) \rangle$$

is measurable.

Proposition A.3 (Weak Measurability of Measure-Valued Maps). *Let X be a compact Hausdorff space and let $\mathcal{M}(X)$ denote the Banach space of finite Radon measures on X , equipped with the weak* topology induced by the duality with $C_b(X)$. Let*

$$\mu : \Omega \rightarrow \mathcal{M}(X)$$

be a map from a measurable space (Ω, \mathcal{F}) . Then the following are equivalent:

1. *The map μ is weakly* measurable, i.e., measurable with respect to the weak* topology $\sigma(\mathcal{M}(X), C_b(X))$.*
2. *For every $f \in C_b(X)$, the function*

$$\omega \mapsto \int_X f(x) d\mu(\omega)(x)$$

is measurable.

Proof. By the Riesz Representation Theorem, $\mathcal{M}(X) \cong C_b(X)^*$, where each $\mu \in \mathcal{M}(X)$ acts on $f \in C_b(X)$ via

$$\langle \mu, f \rangle = \int_X f d\mu.$$

This defines the weak* topology on $\mathcal{M}(X)$, so a map $\mu : \Omega \rightarrow \mathcal{M}(X)$ is weakly* measurable if and only if the scalar function

$$\omega \mapsto \langle \mu(\omega), f \rangle = \int_X f d\mu(\omega)$$

is measurable for every $f \in C_b(X)$. This is exactly statement 2, so the two are equivalent. \square

A.2 Measure

Proposition A.4 (Properties of a Measure [BK13, pg. 20]). *For any measure μ and any measurable sets A, A_1, A_2, \dots , there holds:*

(i) **Monotonicity:** $\mu(A_1) \leq \mu(A_2)$, if $A_1 \subseteq A_2$,

(ii) **σ -subadditivity:**

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n),$$

(iii) **σ -continuity:** If $A_n \uparrow A$, then $\mu(A_n) \rightarrow \mu(A)$ as $n \rightarrow \infty$. If $A_n \downarrow A$ and moreover $\mu(A_1) < \infty$, then $\mu(A_n) \rightarrow \mu(A)$ as $n \rightarrow \infty$ as well.

Definition A.2 (Finite Measure). A measure μ defined on a σ -algebra \mathcal{F} is said to be finite if the measure of the entire space X is finite, that is,

$$\mu(X) < \infty.$$

Definition A.3 (σ -Finite Measure). A measure μ defined on a σ -algebra \mathcal{F} is said to be σ -finite if the space X can be decomposed into a countable union of sets $A_n \in \mathcal{F}$, such that each set A_n has finite measure, i.e.,

$$X = \bigcup_{n=1}^{\infty} A_n \quad \text{and} \quad \mu(A_n) < \infty \text{ for all } n \in \mathbb{N}.$$

Definition A.4. Let (S, \mathcal{A}) , (S', \mathcal{A}') be measurable spaces, let $\varphi : S \rightarrow S'$ be measurable, and let μ be a measure on \mathcal{A} . The measure μ' on S' given by

$$\mu'(A') = \mu(\varphi^{-1}(A')), \quad A' \in \mathcal{A}',$$

is called the image measure of μ under the mapping φ . We write $\mu' = \varphi(\mu) = \mu \circ \varphi^{-1}$ [BK13, pg. 22].

For the next Proposition we need the notion of \cap -stability. We say a generator \mathcal{E} is \cap -stable if $E_1, E_2 \in \mathcal{E} \implies E_1 \cap E_2 \in \mathcal{E}$.

Proposition A.5 (Uniqueness Theorem for Measures, [BK13, Proposition 7.1]). Let \mathcal{E} be a \cap -stable generator of a σ -algebra \mathcal{A} on a set S , and let μ and ν be two measures on \mathcal{A} . If

- (i) $\mu(E) = \nu(E)$ for every $E \in \mathcal{E}$,
- (ii) $\mu(S) = \nu(S) < \infty$ or $\mu(E_n) = \nu(E_n) < \infty$ for some sets $E_1, E_2, \dots \in \mathcal{E}$ with $E_n \uparrow S$,

then $\mu = \nu$.

Definition A.5 (Absolute Continuity). A measure μ on a measurable space (X, \mathcal{A}) is said to be absolutely continuous with respect to a measure λ , written $\mu \ll \lambda$, if for every measurable set $A \in \mathcal{A}$ with $\lambda(A) = 0$, it follows that $\mu(A) = 0$.

Remark A.1. If $\mu \ll \lambda$, and λ is the Lebesgue measure on \mathbb{R}^n , then by the Radon–Nikodym theorem, there exists a measurable function $g \in L^1(\lambda)$ such that

$$\mu(A) = \int_A g d\lambda \quad \text{for all measurable } A \subseteq X.$$

The function $g = \frac{d\mu}{d\lambda}$ is called the Radon–Nikodym derivative or the density of μ with respect to λ .

Let $f : X \rightarrow \mathbb{R}$ be a measurable function. If $\mu \ll \lambda$ and $\frac{d\mu}{d\lambda} = g$, then

$$\int_X f d\mu = \int_X f g d\lambda.$$

If g is continuous, then $f g$ is Lebesgue measurable, and:

- If f is integrable with respect to λ , then $f g \in L^1(\lambda)$, so $\int_X f g d\lambda$ exists.
- If f is continuous and $g \in L^1(\lambda)$, then $f g \in L^1(\lambda)$, since the product of a continuous and an integrable function is integrable on a compact domain (or under appropriate decay conditions).

A.3 Integral

Proposition A.6 (Monotone Convergence Theorem [BK13, pg. 31]). *Let $f_1 \leq f_2 \leq \dots$ hold for measurable functions f_1, f_2, \dots , and set $f = \sup_{n \geq 1} f_n$. Then*

$$\int f d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu$$

Theorem A.1 (Tonelli's Theorem – Integrability Criterion [Bog07a, Thm. 3.4.5]). *Let $f : X \times Y \rightarrow [0, \infty]$ be a nonnegative measurable function with respect to the product σ -algebra, where μ and ν are σ -finite measures on X and Y , respectively. Then:*

$$f \in L^1(\mu \otimes \nu) \quad \text{if and only if} \quad \int_Y \left(\int_X f(x, y) d\mu(x) \right) d\nu(y) < \infty.$$

Moreover, in that case,

$$\int_{X \times Y} f(x, y) d(\mu \otimes \nu)(x, y) = \int_Y \left(\int_X f(x, y) d\mu(x) \right) d\nu(y) = \int_X \left(\int_Y f(x, y) d\nu(y) \right) d\mu(x).$$

B Proofs and Discussions

B.1 Proof of Lemma 1.1

Lemma B.1. *Let $\{\mu_i\}_{i \in \mathbb{N}}$ be a collection of finite measures, where each μ_i is defined on a corresponding σ -algebra \mathcal{A}_i for each $i \in \mathbb{N}$. Then $\mu = \sum_i \mu_i$ is a measure on the σ -algebra $\mathcal{A} = \bigcap_i \mathcal{A}_i$. The measure space (S, \mathcal{A}, μ) enjoys the property that for any measurable function $f : S \rightarrow \overline{\mathbb{R}}$, we have that*

$$\int f d\mu = \sum_i \int f d\mu_i, \tag{B.12}$$

holds. The symbol $\overline{\mathbb{R}}$ denotes the closure of \mathbb{R} . For convenience, and with some abuse of notation, we write $\overline{\mathbb{R}} = [-\infty, \infty]$. Conventions with regard to the symbol ∞ are discussed in Section B.2.

Proof. We shall begin by verifying that μ is indeed a measure on \mathcal{A} . Monotonicity and σ -continuity are immediate from the fact that the individual μ_i 's are measures (cf. Proposition A.4). The σ -subadditivity property follows from the fact that for any disjoint sequence $\{A_n\}_{n \geq 1}$, where $A_n \in \mathcal{A}$ and $\bigcap_i A_i = \emptyset$, we have that

$$\mu(\cup_n A_n) = \sum_i \sum_n \mu_i(A_n) = \sum_n \sum_i \mu_i(A_n) = \sum_n \mu(A_n),$$

where convergence of the sum follows from the finiteness of the μ_i 's and the fact that $A_n \in \mathcal{A}_i$. Next, we show the integration property of the lemma (cf. Eq. (B.12)). For that, let $f \geq 0$ be an elementary function. Then, we have that

$$\int f d\mu = \sum_z z \mu(\{f = z\}) = \sum_z z \sum_i \mu_i(\{f = z\}) = \sum_i \sum_z z \mu_i(\{f = z\}) = \sum_i \int f d\mu_i,$$

where again the convergence of the sum follows from the finiteness of the μ_i 's. Now we shall consider for measurable $f \geq 0$, the sequence $\{f_n\}_{n \in \mathbb{N}}$, with

$$f_n = \sum_k \frac{k}{2^n} I_{\{\frac{k}{2^n} \leq f \leq \frac{k+1}{2^n}\}} + 2^n I_{\{f = \infty\}},$$

having the properties that $f_1 \leq f_2 \dots$ and $\sup_{n \geq 1} f_n = f$. This exact same construction was used in [BK13, pg.34]. In fact it can be shown that for any measurable function f there is a sequence of elementary functions, with the desired properties see for example [Bog07a, Corollary 2.1.9]. We then obtain, by Proposition A.6, that

$$\int f d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu = \lim_{n \rightarrow \infty} \sum_i \int f_n d\mu_i = \sum_i \lim_{n \rightarrow \infty} \int f_n d\mu_i,$$

holds. The exchange of the sum and limit is also a consequence of Proposition A.6. Here, the sum is considered as an integral with respect to the counting measure (over \mathbb{N}), defining the sequence $\{g_n\}_{n \in \mathbb{N}}$, with $g_n(i) = \int f_n d\mu_i$, we see that the conditions of Proposition A.6 are satisfied. Explicitly, we have that $g_1 \leq g_2 \leq \dots$ and $\sup_{n \geq 1} g_n(i) = \int f d\mu_i$. Finally, the assertion follows from the fact, that for any integrable function f , we can write $\int f d\mu = \int \max(0, f) d\mu - \int \max(-f, 0) d\mu$. \square

B.2 The extended Real Line $\overline{\mathbb{R}}$

The extension $\overline{\mathbb{R}} = [-\infty, \infty]$, i.e the closure of \mathbb{R} is discussed in [BK13, pg. 13]. The following conventions for the symbol ∞ were made:

- $\infty + \infty := \infty$
- $= 0 \cdot \infty := 0$
- $a \cdot \infty := \infty, a > 0$
- $(-1) \cdot \infty := -\infty$

The expressions $\infty - \infty$ and $\frac{\infty}{\infty}$, however remain undefined.

B.3 Discussion on Metric Spaces which are the image of a continuous f under compact interval

For a discussion about compact metric spaces, which are the image of a continuous function over the unit interval $[0, 1]$, the following theorem is really helpful:

Theorem B.1 (Hahn–Mazurkiewicz theorem [JGH12, Theorem 3.30]). *A non-empty Hausdorff topological space is a continuous image of the unit interval if and only if it is a compact, connected, locally connected, second-countable space.*

We do not offer a rigorous definition of second countable, however we mention that a compact metric space, by definition, is second countable. Similar for connect/ locally connected property, we state without proof that the compact metric space $[0, 1]^n, d_{[0,1]^n}$ possesses these properties for all choices of n and all metrics $d_{[0,1]^n}$. We end the discussion, by stating the obvious fact that continuous maps generating the metric spaces are in general highly non smooth.

B.4 Comparison with the ergodicity metric as originally defined in [MM21]

We shall now relate the ergodic framework developed to the metric introduced in [MM21], by [MM21, Eq. (1)] the metric is given by:

$$d^t(B(x, r)) = \frac{1}{Nt} \sum_{j=1}^N \int_0^t I_{B(x, r)}(g_j(\tau)) \lambda(d\tau),^3 \quad (\text{B.13})$$

where $B(x, r)$ denotes the spherical subsets of U centered at x , formally $B(x, r) = \{y \in U \mid \|y - x\|^2 \leq r\}$. Following Eq. (B.13) we define the system of subsets $\mathcal{E} = \{B(x, r) \mid x \in U, r > 0\}$, where $B(x, r)$ was already defined before. We state that for a compact domain $U \subset \mathbb{R}^n$ any open set can be written as a union of the spherical sets in U and by extension \mathcal{E} generates the Borel σ -algebra of U . At this point we state that all trajectories $g_j : [0, t] \rightarrow U$, as they are continuous functions, are $\mathcal{B} - \mathcal{B}^n$ measurable, this is a direct consequence of Proposition A.2, as for continuous functions, the preimages of open sets are again open and hence Borel-measurable. Finally, extending Eq. (B.13) to the Borel sets in U , by considering Borel sets instead of spherical sets, we get the following equation:

$$d^t(B) = \frac{1}{Nt} \sum_{j=1}^N \int_0^t I_B(g_j(\tau)) \lambda(d\tau), B \in \mathcal{B}^n. \quad (\text{B.14})$$

³For better readability we altered the original form, given by: $d^t(x, r) = \frac{1}{Nt} \sum_{j=1}^N \int_0^t \chi(x, r)(x_j(\tau)) d\tau$

Using Lemma 3.1 and 1.1, together with the fact that the probability measure of the uniform distribution over some interval $[0, t]$ is given by $\frac{\lambda(\tau)}{\lambda([0, t])}$, $\tau \subseteq [0, t]$, measurable, we see that in fact the left hand side of Eq. (1.2) and Eq. (B.14) are in fact equivalent. Finally, as stated before all spherical subset of U are Borel-measurable, therefore Eq. (B.14) is an extension of (B.13).

References

- [Bax10] Brad Baxter. Conditionally positive functions and p-norm distance matrices, 2010.
- [BK13] Martin Brokate and Götz Kersting. *Measure and Integral*. Springer, Berlin, 2013.
- [Bog07a] V. I. Bogachev. *Measure Theory, Volume I*. Springer-Verlag, Berlin, 2007.
- [Bog07b] V. I. Bogachev. *Measure Theory, Volume II*. Springer-Verlag, Berlin, 2007.
- [GS64] I. M. Gel’fand and G. E. Shilov. *Generalized Functions, Volume 1: Properties and Operations*. Academic Press, 1964.
- [JGH12] Gail S. Young John G. Hocking. *Topology*. Courier Corporation, 2012.
- [Mil76] Stephen C. Milne. Peano curves and smoothness of functions. *Proceedings of the American Mathematical Society*, 58(1):167–173, 1976.
- [MM13] Lauren M. Miller and Todd D. Murphey. Trajectory optimization for continuous ergodic exploration on the motion group $se(2)$. In *52nd IEEE Conference on Decision and Control*, pages 4517–4522, 2013.
- [MM21] George Mathew and Igor Mezić. Metrics for ergodicity and design of ergodic dynamics for multi-agent systems. 2021.
- [Ope25] OpenAI. ChatGPT: Language Model (May 2025 Version). <https://chat.openai.com>, 2025. Generative AI model developed by OpenAI. Available at: <https://chat.openai.com>.
- [Rud87] Walter Rudin. *Real and Complex Analysis*. McGraw-Hill, New York, 3rd edition, 1987.
- [Rud91] Walter Rudin. *Functional Analysis*. McGraw-Hill, New York, 2nd edition, 1991.
- [Sag94] Hans Sagan. *Space-Filling Curves*. Springer-Verlag, New York, 1994.
- [SGTM25] Max Muchen Sun, Ayush Gaggar, Pete Trautman, and Todd Murphey. Fast ergodic search with kernel functions. *IEEE Transactions on Robotics*, 41:1841–1860, 2025.