

IMPRECISE MARKOV SEMIGROUPS AND THEIR ERGODICITY

MICHELE CAPRIO^{1,*}  AND MENGQI CHEN²

Abstract. We introduce the concept of an imprecise Markov semigroup $\mathbf{Q}$, a tool that represents ambiguity around both the transition probabilities and the invariant measure of a continuous-time Markov process *via* a collection of Markov semigroups, each associated with a (possibly different) Markov process. We use techniques from topology, geometry, and probability to analyze ergodic limits under model uncertainty encoded by $\mathbf{Q}$. We establish long-term bounds that are uniform in the initial state and identify regimes in which the imprecision in these bounds collapses asymptotically. Our results are proved in progressively more general settings. We first assume that $\mathbf{Q}$ is compact and that the state space is Euclidean or a Riemannian manifold, working with a fixed bounded observable. We then allow the state space to be standard Borel, while keeping $\mathbf{Q}$ compact and the observable fixed. Finally, we drop compactness and work on Polish metric spaces of finite diameter, where we treat arbitrary bounded Lipschitz observables. The importance of our findings for the fields of artificial intelligence and computer vision is also discussed at a high level; In particular, in the study of how the probability of an output evolves over time as we perturb the input of a convolutional autoencoder.

Mathematics Subject Classification. 60J60, 58J65, 62A01.

Received March 7, 2026. Accepted May 19, 2026.

1. INTRODUCTION

When dealing with a stochastic phenomenon whose future evolution is independent of its history, it is customary to model its behavior *via* a stochastic process enjoying the Markov property. Loosely, this means that the law of the realization X_T of the process $(X_t)_{t \geq 0}$ at time T , given what happened up to time $s < T$, is equivalent to the law of X_T , given only X_s . We can of course express this property in a slightly more general way, by saying that the expectation of a generic bounded measurable real-valued function $\tilde{f}$ of X_T , given what happened in the past up to time $s < T$, is equal to the expectation of $\tilde{f}(X_T)$, given only X_s . Notice that in the present paper we perpetrate a terminology abuse, and refer to real-valued functions on the state space E as “functionals”. In addition, we only consider continuous-time processes $(X_t)_{t \geq 0}$.

Two elements play a key role in a process enjoying the Markov property. The first is the law of the initial state X_0 , and the second is the transition probability from the initial state X_0 to any state X_t of the process, *i.e.* the probability that, given that the process begins at X_0 , it “reaches a certain value” at time t . In many

Keywords and phrases: Imprecise Markov processes, Markov diffusion operators, Bakry–Émery curvature, ergodic processes, autoencoders.

¹ Department of Computer Science, The University of Manchester, Manchester M13 9PL, UK.

² Department of Statistics, University of Warwick, Coventry CV4 7AL, UK.

* Corresponding author: michele.caprio@manchester.ac.uk

applications, relaxing the hypothesis that one or both of them are precise is natural, and leads to less precise but more robust results.

Imprecise probability [1–3] is a field whose techniques naturally lend themselves to be used for this purpose. In fact, imprecise Markov processes are a flourishing area of the imprecise probability literature [4–14],¹ with numerous applications *e.g.* in engineering [15], Chapter 4, including aerospace engineering [16], Chapter 5, and economics [17–19]. As the name suggests, imprecise Markov processes allow to take into account imprecision in the initial and transition probabilities (and also in the invariant measure, when it exists) of a Markov process.

In this paper, we introduce a new concept that we call an *Imprecise Markov Semigroup*. It is a collection $\mathbf{Q}$ of (precise) Markov semigroups, each associated with a Markov process.² Modeling the imprecision on the transition probabilities and on the invariant measure (whenever it exists) *via* such a collection, instead of using lower previsions [2] or Choquet integrals [20], allows us to combine geometric and probabilistic techniques based on the Markov diffusion operators literature [32] that make the study of the limiting behavior of imprecise Markov processes (captured by the limit behavior of the imprecise Markov semigroup) easier. In particular, we are able to relate such a behavior to the curvature of the state space – that is, the space E that the elements of the process $(X_t)_{t \geq 0}$ take values on. This also links the present work to the study of the geometry of uncertainty [21].

Our goal is to study what we call the *ergodic property* of an imprecise Markov semigroup. Throughout the paper we use “ergodicity” in the semigroup/long-term sense: We study the asymptotic behavior, as $t \rightarrow \infty$, of conditional expectations of observables under the Markov evolution, *i.e.* $P_t f(x) = \mathbb{E}[f(X_t) \mid X_0 = x]$. In Sections 3 and 4, the bounds we derive are formulated for a *fixed* (bounded) test functional f ; Accordingly, the conclusions should be read as long-term bounds for the expectation of $\tilde{f}$ whose limiting constants do not depend on the initial state, rather than as full distributional convergence – which would correspond to convergence for all $f \in B(E)$. In contrast, in Section 5 we work on the larger class $\text{Lip}_b(E)$ of bounded Lipschitz functionals and obtain uniform-in- x limits for the static upper and lower envelopes $\overline{T}_t$ and $\underline{T}_t$ of $\mathbf{Q}$, respectively, and for the time-consistent (upper/lower) Nisio semigroups when switching is allowed. Such results are derived for every bounded Lipschitz observable f , which is closer in spirit to distributional stability. Finally, we emphasize that our limits are *not* Cesàro time averages: We do *not* prove statements of the form $\frac{1}{T} \int_0^T f(X_t) dt \rightarrow \int f d\mu$. Rather, we establish convergence of the *semigroup* as $t \rightarrow \infty$ (equivalently, convergence of one-time marginals and/or of their upper/lower envelopes when switching is allowed). When appropriate (*e.g.* under standard invariance/ergodicity assumptions), classical ergodic theorems can then be invoked to relate such semigroup convergence to time average statements, but this is not the object of the present results.

The similarities and differences between our results and the existing ones in the imprecise Markov literature and in imprecise ergodic theory [22–29], are discussed in Sections 2 and 3, and 5.

Besides being interesting in their own right (and because they provide new techniques to the fields of imprecise Markov processes), we believe that our findings will prove useful in future research in machine learning, artificial intelligence, and computer vision. Let us add a high-level discussion on this topic. In recent works on convolutional autoencoders, images (or, more in general, inputs from an input space $\mathcal{X}$) are “projected” onto a low-dimensional manifold E by an encoding function $\phi_{\text{enc}} : \mathcal{X} \rightarrow E$. Different “portions” of the surface $\mathcal{S}(E)$ of such a manifold capture different features of the input image [30]. For instance, a portion of the surface may correspond to the feature “dog”, so that dog images are projected there. This means that we can consider a partition $\mathcal{E} = \{E_j\}_{j=1}^J$ of $\mathcal{S}(E)$ whose elements correspond to “feature portions”. For example, J could be equal to the cardinality $|\mathcal{Y}|$ of the output space $\mathcal{Y}$ in the case of classification problems. Then, a decoding function $\phi_{\text{dec}} : E \rightarrow \Delta_{\mathcal{Y}}$ returns the probabilities of $y_j \in \mathcal{Y}$ being the correct output for the input $x \in \mathcal{X}$. Here $\Delta_{\mathcal{Y}}$ denotes the space of countably additive probabilities on $\mathcal{Y}$. For example, in a classification case, $(\phi_{\text{dec}} \circ \phi_{\text{enc}})(x) = (P(y_1), \dots, P(y_J))^{\top}$, where $\mathcal{Y} = \{y_1, \dots, y_J\}$ is the label space, and $P(y_j)$ is the probability estimated by the model that y_j is the correct label for input x , $j \in \{1, \dots, J\}$. A visual representation is given in Figure 1.

¹Much of the existing literature focuses on discrete-time processes.

²Here, by “precise” we refer to the classical notion of a Markov semigroup.

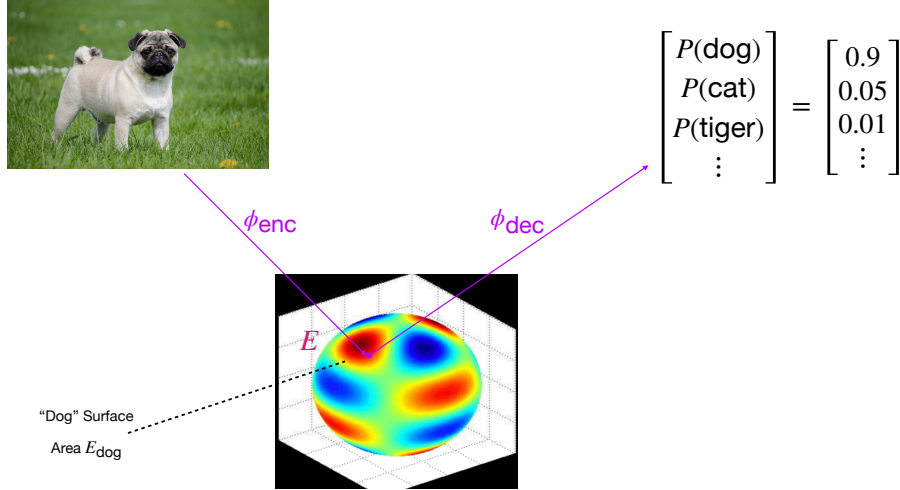


FIGURE 1. In this figure, we provide a visual representation of the functioning of a convolutional autoencoder in an image classification problem. The input image of a pug is mapped by encoding function ϕ_{enc} to the surface area $E_{\text{dog}} \in \mathcal{E}$ corresponding to the feature “dog”. Such a projection is then mapped by decoding function ϕ_{dec} to the probability space over the labels. In this simple example the model gives a high probability to the first label, *i.e.* “dog”.

Now, imagine that we want to study a smooth transition between a dog and a cat image. Intuitively, this corresponds to a “walk” from the “dog portion” E_{dog} of the manifold’s surface $\mathcal{S}(E)$ to the “cat portion” E_{cat} [31]. A visual example is given in Figure 2. Our results, and especially Corollaries 3.3 and 5.10, may be very useful when (i) such a “walk” is not deterministic, but rather modeled *via* a Markov process; (ii) the user is uncertain about which transition probabilities to choose (and on the nature of the invariant measure); and (iii) we are interested in the dynamics of the probability of a particular label during such a walk. For example, if we are interested in how the probability of the first label evolves over time, we could consider function $z \mapsto \tilde{f}(z) = e_1^\top \phi_{\text{dec}}(z)$, where $e_1 = (1, 0, 0, \dots, 0)^\top$ is a J -dimensional standard basis vector, and $\tilde{f}$ is an example of a bounded functional on E , as discussed earlier in this section.

Our paper is structured as follows. Section 2 introduces the concept of an Imprecise Markov Semigroup $\mathbf{Q}$. Section 3 derives our results when state space E is a Euclidean space or a Riemannian manifold, $\mathbf{Q}$ is compact, and the bounded test functional $\tilde{f}$ is fixed. Section 4 generalizes the findings in Section 3 to the case of E being an arbitrary measurable space, while keeping $\mathbf{Q}$ compact and the bounded test functional $\tilde{f}$ fixed. Section 5 further extends our results: It is only required that E is Polish metric with finite diameter and that the test map is any bounded Lipschitz functional. Section 6 concludes our work. We prove Theorem 4.2 in Appendix A, and we provide a primer on Markov semigroups and diffusion Markov triples in Appendix C.

2. IMPRECISE MARKOV SEMIGROUPS

In this section, we introduce the concept of an Imprecise Markov Semigroup. Its definition hinges on the notions of Markov processes, Markov semigroups, carré du champ and iterated carré du champ operators, and Bakry-Émery curvature. We refer the unfamiliar reader to Appendix C.1.

The following is introduced in [32], Chapter 1.1. It is the minimal structure we require throughout the paper for the space where our stochastic processes of interest live.

Definition 2.1 (Good Measurable Space [32]). A *good measurable space* is a measurable space $(E, \mathcal{F})$ for which the measure decomposition theorem applies, and for which there is a countable family of sets which generates the σ -algebra $\mathcal{F}$.

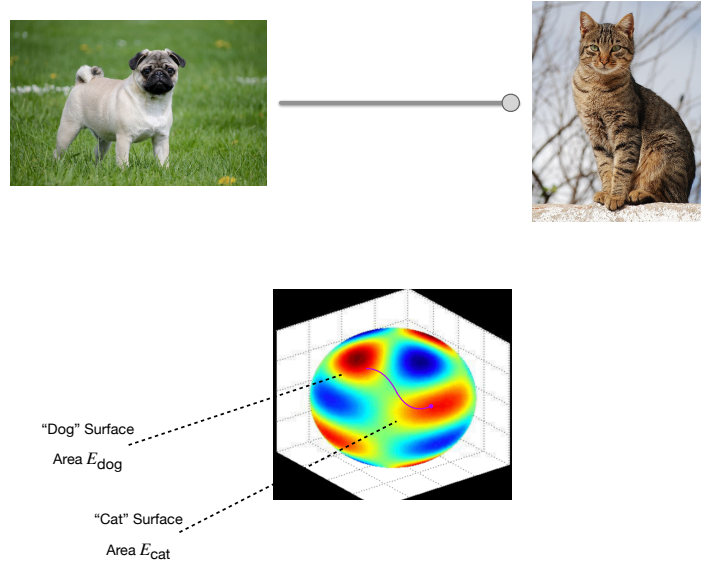


FIGURE 2. In this figure, we represent the smooth transition between a dog and a cat pictures. Moving a cursor between the pictures causes a walk from the “dog portion” E_{dog} of the manifold’s surface $\mathcal{S}(E)$ to the “cat portion” E_{cat} , depicted as a purple smooth curve. The results we present in this paper are useful to determine the behavior of a similar walk that is not deterministic, when it is not possible to determine exactly transition probabilities (and the invariant measure, when it exists), and when we are interested *e.g.* in the evolution of the probability of the first label, $y_1 = \text{“dog”}$, captured by the functional $z \mapsto \tilde{f}(z) = e_1^\top \phi_{\text{dec}}(z)$.

Here, the measure decomposition theorem refers to the *disintegration theorem* (existence of regular conditional distributions). Concretely, if μ is a probability measure on $(E \times E, \mathcal{F} \otimes \mathcal{F})$ and μ_1 is its first marginal, then there exists a Markov kernel $k(x, dy)$ (measurable in x) such that $\mu(dx, dy) = k(x, dy)\mu_1(dx)$. Equivalently, μ decomposes into the conditional laws of the second coordinate given the first. This is typically guaranteed on standard Borel spaces (in particular, Polish spaces with their Borel σ -algebras).

In this work, we are interested in Markov semigroups $\mathbf{P} = (P_t)_{t \geq 0}$ associated with (time homogeneous) Markov processes $(X_t)_{t \geq 0}$. For such a Markov semigroup, for every $t \geq 0$, operator P_t is such that

$$P_t f(x) := \mathbb{E}[f(X_t) \mid X_0 = x] = \int_E f(y) p_t(x, dy), \quad f \in B(E), x \in E,$$

where $B(E)$ is the space of bounded measurable functionals on E , and p_t is the probability kernel describing the evolution of the Markov process. In words, operator P_t applied on $f \in B(E)$ gives us the expected value of $f(X_t)$, “weighted by” the transition probability (from the starting value $X_0 = x$ to X_t) of our Markov process. We focus on this type of Markov semigroups because studying their (ergodic) behavior informs us about the (ergodic) behavior of the associated Markov processes.

Throughout the paper, we will assume that the semigroups we consider always have an invariant probability measure μ , so that the elements of $\mathbf{P} = (P_t)_{t \geq 0}$ are bounded operators in $\mathbb{L}^p(\mu)$, $p \in [1, \infty)$. In fact, most semigroups of interest have an invariant measure [32], Section 1.2.1. A way to derive μ as a weak limit of a “reasonable” initial probability measure μ_0 is inspected in [32], p. 10. For a discussion on this matter, we refer the interested reader to Remark C.2 in Appendix C.1.

We now introduce a space that will play a pivotal role in the definition of an Imprecise Markov Semigroup. Recall that, for all $t \geq 0$, the t -th element P_t of Markov semigroup $\mathbf{P} = (P_t)_{t \geq 0}$ is a linear operator that sends bounded measurable functionals on E to bounded measurable functionals on E . In formulas, we can write this as $P_t : B(E) \rightarrow B(E)$, or equivalently, $P_t \in B(E)^{B(E)}$. In turn, this implies that a Markov semigroup $\mathbf{P}$ can be seen as a function mapping bounded measurable functionals f on E to nets of bounded measurable functionals $\mathbf{P}f = (P_t f)_{t \geq 0}$ on E . In formulas, $\mathbf{P} : B(E) \rightarrow B(E)^{\mathbb{R}_+}$, or equivalently, $\mathbf{P} \in (B(E)^{\mathbb{R}_+})^{B(E)}$.

Definition 2.2 (Imprecise Markov Semigroup (IMSG)). An Imprecise Markov Semigroup is a nonempty collection $\mathbf{Q}$ of Markov semigroups (MSGs) associated with Markov processes. That is, $\mathbf{Q} \subset (B(E)^{\mathbb{R}_+})^{B(E)}$ such that

$$\mathbf{Q} \subseteq \mathfrak{Q} := \{\mathbf{P} \in (B(E)^{\mathbb{R}_+})^{B(E)} : \mathbf{P} \text{ is an MSG associated with a Markov process}\}. \quad (2.1)$$

In Sections 3 and 4, we will work with compact IMSGs. To do so, we need to introduce a notion of compactness in the space $(B(E)^{\mathbb{R}_+})^{B(E)}$. Such a notion hinges on the possibility of comparing the elements of $\mathbf{Q}$, when they are evaluated at some functional $\tilde{f} \in B(E)$.

First, suppose we are able to define a total preorder $\preceq_{\tilde{f}}^{\text{tot}}$ so that

$$\mathbf{P} \preceq_{\tilde{f}}^{\text{tot}} \mathbf{P}' \iff P_t \tilde{f} \leq P'_t \tilde{f}, \forall t \geq 0.$$

This relation is always reflexive and transitive on $\mathbf{Q}$; It need not be antisymmetric, since distinct semigroups may coincide on $\tilde{f}$. We therefore introduce the equivalence relation

$$\mathbf{P} \sim_{\tilde{f}} \mathbf{P}' \iff P_t \tilde{f} = P'_t \tilde{f} \text{ for all } t \geq 0,$$

and work on the quotient set $\mathbf{Q}_{\tilde{f}} := \mathbf{Q} / \sim_{\tilde{f}}$. The preorder $\preceq_{\tilde{f}}^{\text{tot}}$ induces a genuine *total order* on $\mathbf{Q}_{\tilde{f}}$ by

$$[\mathbf{P}] \preceq_{\tilde{f}}^{\text{tot}} [\mathbf{P}'] \iff \mathbf{P} \preceq_{\tilde{f}}^{\text{tot}} \mathbf{P}',$$

which is well-defined and antisymmetric by construction. Of course, here the square brackets denote equivalence classes. We have that, for all $[\mathbf{P}], [\mathbf{P}'] \in \mathbf{Q}_{\tilde{f}}$, either $[\mathbf{P}] \preceq_{\tilde{f}}^{\text{tot}} [\mathbf{P}']$ or $[\mathbf{P}'] \preceq_{\tilde{f}}^{\text{tot}} [\mathbf{P}]$. In addition, the notions of maximal and greatest elements of $(\mathbf{Q}_{\tilde{f}}, \preceq_{\tilde{f}}^{\text{tot}})$ coincide, and correspond to the classical notion of maximum (with respect to the total order $\preceq_{\tilde{f}}^{\text{tot}}$). Similarly, minimal and least element notions coincide, and correspond to the classical notion of minimum.

Now, let $\tau_{\preceq_{\tilde{f}}^{\text{tot}}}$ be the order topology on $\mathbf{Q}_{\tilde{f}}$. That is (provided $|\mathbf{Q}_{\tilde{f}}| \geq 2$), $\tau_{\preceq_{\tilde{f}}^{\text{tot}}}$ is the topology whose base is given by the open rays $\{[\mathbf{P}] : [\mathbf{P}_1] \prec_{\tilde{f}}^{\text{tot}} [\mathbf{P}]\}$ and $\{[\mathbf{P}] : [\mathbf{P}] \prec_{\tilde{f}}^{\text{tot}} [\mathbf{P}_2]\}$, where $\prec_{\tilde{f}}^{\text{tot}}$ is the strict counterpart of $\preceq_{\tilde{f}}^{\text{tot}}$, for all $[\mathbf{P}_1], [\mathbf{P}_2]$ in $\mathbf{Q}_{\tilde{f}}$, and by the open intervals $([\mathbf{P}_1], [\mathbf{P}_2]) = \{[\mathbf{P}] : [\mathbf{P}_1] \prec_{\tilde{f}}^{\text{tot}} [\mathbf{P}] \prec_{\tilde{f}}^{\text{tot}} [\mathbf{P}_2]\}$. This means that the $\tau_{\preceq_{\tilde{f}}^{\text{tot}}}$ -open sets in $\mathbf{Q}_{\tilde{f}}$ are the sets that are a union of (possibly infinitely many) such open intervals and rays.

Assume that $\mathbf{Q}_{\tilde{f}}$ is $\tau_{\preceq_{\tilde{f}}^{\text{tot}}}$ -compact. This implies that $\mathbf{Q}_{\tilde{f}}$ contains the minimum and the maximum (with respect to the total order $\preceq_{\tilde{f}}^{\text{tot}}$), which we denote by $[\underline{\mathbf{P}}]_{\tilde{f}}$ and $[\overline{\mathbf{P}}]_{\tilde{f}}$, respectively. We fix arbitrary representatives $\underline{\mathbf{P}}_{\tilde{f}} \in [\underline{\mathbf{P}}]_{\tilde{f}}$ and $\overline{\mathbf{P}}_{\tilde{f}} \in [\overline{\mathbf{P}}]_{\tilde{f}}$. In turn, since $\underline{\mathbf{P}}_{\tilde{f}}, \overline{\mathbf{P}}_{\tilde{f}} \in \mathbf{Q}$, they both are (precise) Markov semigroups associated with Markov processes by (2.1).

Suppose now that the induced relation is not total. More formally, the preorder $\preceq_{\tilde{f}}^{\text{part}}$ on $\mathbf{Q}$ is defined similarly to before by

$$\mathbf{P} \preceq_{\tilde{f}}^{\text{part}} \mathbf{P}' \iff P_t \tilde{f} \leq P'_t \tilde{f}, \forall t \geq 0.$$

Consider its induced relation (a genuine partial order) on the quotient $\mathbf{Q}_{\tilde{f}}$. Contrary to before, this time there might exist $[\mathbf{P}], [\mathbf{P}'] \in \mathbf{Q}_{\tilde{f}}$ such that $[\mathbf{P}] \not\preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}']$ and $[\mathbf{P}] \not\preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}']$. A remark is in order.

Remark 2.3 (Antisymmetry and Identifiability). Fix $\tilde{f} \in B(E)$ and consider the preorder on $\mathbf{Q}$ defined by

$$\mathbf{P} \preceq_{\tilde{f}} \mathbf{P}' \iff P_t \tilde{f} \leq P'_t \tilde{f}, \forall t \geq 0.$$

This preorder is *antisymmetric* on $\mathbf{Q}$ (hence a genuine partial order on the semigroups in $\mathbf{Q}$) if and only if the orbit map

$$\mathbf{P} \mapsto (P_t \tilde{f})_{t \geq 0} \in B(E)^{\mathbb{R}_+}$$

is injective on $\mathbf{Q}$, *i.e.* whenever $P_t \tilde{f} = P'_t \tilde{f}$ for all $t \geq 0$, it follows that $\mathbf{P} = \mathbf{P}'$ as semigroups. Equivalently, $\tilde{f}$ *separates* the elements of $\mathbf{Q}$ through their time evolution. In general, a single functional $\tilde{f}$ will not be separating, so assuming antisymmetry from $\tilde{f}$ can be a strong identifiability restriction on $\mathbf{Q}$. However, antisymmetry can hold in structured families (*e.g.* low-dimensional parametric classes) when $\tilde{f}$ is chosen so that the trajectories $t \mapsto P_t \tilde{f}$ uniquely identify the underlying semigroup within $\mathbf{Q}$. Identifiability of Markov kernel/semigroup can be guaranteed *via* a *separating (measure-determining) class* of test functions rather than on a single $\tilde{f}$ [33, 34]. In this paper we instead allow an arbitrary, application-driven choice of $\tilde{f}$ and handle potential non-antisymmetry by working with the quotient $\mathbf{Q}_{\tilde{f}} := \mathbf{Q} / \sim_{\tilde{f}}$. In the special case where $\tilde{f}$ is separating on $\mathbf{Q}$, the equivalence relation $\sim_{\tilde{f}}$ is trivial and the quotient poset becomes unnecessary.

Let us now introduce three set theoretic concepts, inspired by [35], that will allow us to properly define a compact Imprecise Markov Semigroup even when only a partial order can be determined.

Definition 2.4 (Up/Down-Directed Subsets of $\mathbf{Q}_{\tilde{f}}$). A subset $\mathbf{S}$ of $\mathbf{Q}_{\tilde{f}}$ is *up-directed* if for all $[\mathbf{P}], [\mathbf{P}'] \in \mathbf{S}$, there exists $[\mathbf{P}'''] \in \mathbf{S}$ such that $[\mathbf{P}] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}''']$ and $[\mathbf{P}'] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}''']$. We say that $\mathbf{S}$ is *down-directed* if the opposite holds. That is, if for all $[\mathbf{P}], [\mathbf{P}'] \in \mathbf{S}$, there exists $[\mathbf{P}'''] \in \mathbf{S}$ such that $[\mathbf{P}'''] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}]$ and $[\mathbf{P}'''] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}']$.

Definition 2.5 (Dedekind-Closed Subsets of $\mathbf{Q}_{\tilde{f}}$). A subset $\mathbf{K}$ of $\mathbf{Q}_{\tilde{f}}$ is *Dedekind-closed* if whenever $\mathbf{S}$ is an up-directed subset of $\mathbf{K}$, and $[\mathbf{P}]$ is the $\preceq_{\tilde{f}}^{\text{part}}$ -least upper bound of $\mathbf{S}$ (or $\mathbf{S}$ is a down-directed subset of $\mathbf{K}$, and $[\mathbf{P}]$ is the $\preceq_{\tilde{f}}^{\text{part}}$ -greatest lower bound of $\mathbf{S}$), then $[\mathbf{P}] \in \mathbf{K}$.

Definition 2.6 (Incomparability, Diversity, and $\preceq_{\tilde{f}}^{\text{part}}$ -Width of $\mathbf{Q}_{\tilde{f}}$). Two elements $[\mathbf{P}], [\mathbf{P}'] \in \mathbf{Q}_{\tilde{f}}$ are *incomparable* if $[\mathbf{P}] \not\preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}']$ and $[\mathbf{P}] \not\preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}']$. A subset $\mathbf{S}$ of $\mathbf{Q}_{\tilde{f}}$ is *diverse* if $[\mathbf{P}], [\mathbf{P}'] \in \mathbf{S}$ and $[\mathbf{P}] \neq [\mathbf{P}']$ implies that $[\mathbf{P}]$ and $[\mathbf{P}']$ are incomparable. Finally, we define the $\preceq_{\tilde{f}}^{\text{part}}$ -width of $\mathbf{Q}_{\tilde{f}}$ to be the supremum of the set $\{k : k \text{ is the cardinality of a diverse subset of } \mathbf{Q}_{\tilde{f}}\}$.

Definitions 2.4, 2.5, and 2.6 are needed to introduce the concept of $\preceq_{\tilde{f}}^{\text{part}}$ -compatible topology. Requiring that $\mathbf{Q}_{\tilde{f}}$ is compact in such a topology (together with other assumptions) will ensure us that it possesses least and greatest elements (with respect to the partial order $\preceq_{\tilde{f}}^{\text{part}}$).

Definition 2.7 ($\preceq_{\tilde{f}}^{\text{part}}$ -Compatible Topology [35]). A topology $\tau_{\preceq_{\tilde{f}}^{\text{part}}}$ on $\mathbf{Q}_{\tilde{f}}$ is $\preceq_{\tilde{f}}^{\text{part}}$ -compatible if

- every $\tau_{\preceq_{\tilde{f}}^{\text{part}}}$ -closed set is also Dedekind-closed
- every set of the form $\{[\mathbf{P}] \in \mathbf{Q}_{\tilde{f}} : [\mathbf{P}'] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}']\}$ is $\tau_{\preceq_{\tilde{f}}^{\text{part}}}$ -closed.

The next remark explores the connections between the (generic) partial-order-compatible and the interval topologies.

Remark 2.8 (Interval Topology vs $\preceq$ -Compatible Topologies). Let $(X, \preceq)$ be a partially ordered set. [35] considers two topologies naturally associated with the order.

(1) **Interval topology.** The *interval topology* $\mathcal{T}_I$ is the topology for which the family of order intervals

$$[a, b] := \{x \in X : a \preceq x \preceq b\}, \quad a, b \in X,$$

forms a *closed subbasis*. Equivalently, $\mathcal{T}_I$ is the coarsest topology on X for which every interval $[a, b]$ is closed.

(2) **$\preceq$ -compatible topology.** A topology $\mathcal{T}$ on X is said to be $\preceq$ -compatible if: (i) every $\mathcal{T}$ -closed set is Dedekind-closed, and (ii) every interval $[a, b]$ is $\mathcal{T}$ -closed.

Relationship. Condition (ii) implies that any $\preceq$ -compatible topology must contain the interval topology,

$$\mathcal{T}_I \subseteq \mathcal{T}.$$

Moreover, if $\mathcal{T}_D$ denotes the topology whose closed sets are exactly the Dedekind-closed subsets of X , then condition (i) yields the upper bound

$$\mathcal{T} \subseteq \mathcal{T}_D.$$

Hence, $\preceq$ -compatible topologies are precisely those lying between $\mathcal{T}_I$ and $\mathcal{T}_D$,

$$\mathcal{T}_I \subseteq \mathcal{T} \subseteq \mathcal{T}_D.$$

In particular, $\mathcal{T}_I$ is the *minimal* $\preceq$ -compatible topology requirement (it enforces closed intervals only), whereas $\mathcal{T}_D$ is the *maximal* one (it enforces all Dedekind-closed sets to be closed). In special situations these two coincide (e.g. under finite $\preceq$ -width, similarly to Lem. 2.9), yielding uniqueness of the $\preceq$ -compatible topology.

The following lemma summarizes some useful topological (and order-theoretic) properties of $\mathbf{Q}_{\tilde{f}}$.

Lemma 2.9 (Topological Properties of $\mathbf{Q}_{\tilde{f}}$). *If $\mathbf{Q}_{\tilde{f}}$ has finite $\preceq_{\tilde{f}}^{\text{part}}$ -width, then $\preceq_{\tilde{f}}^{\text{part}}$ -compatible topology $\tau_{\preceq_{\tilde{f}}^{\text{part}}}$ is unique. If in addition (a) $\mathbf{Q}_{\tilde{f}}$ is $\tau_{\preceq_{\tilde{f}}^{\text{part}}}$ -compact, (b) $\preceq_{\tilde{f}}^{\text{part}}$ has closed graph, that is, the set*

$$\text{Graph}(\preceq_{\tilde{f}}^{\text{part}}) := \{([\mathbf{P}], [\mathbf{P}']) \in \mathbf{Q}_{\tilde{f}} \times \mathbf{Q}_{\tilde{f}} : [\mathbf{P}] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}']\}$$

is closed in the product topology $\tau_{\preceq_{\tilde{f}}^{\text{part}}} \times \tau_{\preceq_{\tilde{f}}^{\text{part}}}$, and (c) $\mathbf{Q}_{\tilde{f}}$ is both up- and down- directed, then $\mathbf{Q}_{\tilde{f}}$ admits $\preceq_{\tilde{f}}^{\text{part}}$ -least and $\preceq_{\tilde{f}}^{\text{part}}$ -greatest elements. We denote them by $[\underline{\mathbf{P}}]_{\tilde{f}}$ and $[\overline{\mathbf{P}}]_{\tilde{f}}$, respectively.

Proof. The first part of the lemma is a consequence of [35], theorem, p. 528, which also shows that $(\mathbf{Q}_{\tilde{f}}, \tau_{\preceq_{\tilde{f}}^{\text{part}}})$ is Hausdorff.

For the second claim, assume that $\mathbf{Q}_{\tilde{f}}$ is $\tau_{\preceq_{\tilde{f}}^{\text{part}}}$ -compact and that $\preceq_{\tilde{f}}^{\text{part}}$ has closed graph. Consider the family

$$\mathcal{C} := \{\downarrow [\mathbf{P}] : [\mathbf{P}] \in \mathbf{Q}_{\tilde{f}}\}, \quad \downarrow [\mathbf{P}] := \{[\mathbf{R}] \in \mathbf{Q}_{\tilde{f}} : [\mathbf{R}] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}]\}.$$

Each $\downarrow[\mathbf{P}]$ is nonempty and closed: Indeed,

$$\downarrow[\mathbf{P}] = \{[\mathbf{R}] \in \mathbf{Q}_{\tilde{f}} : ([\mathbf{R}], [\mathbf{P}]) \in \text{Graph}(\preceq_{\tilde{f}}^{\text{part}})\},$$

so it is the preimage of the closed set $\text{Graph}(\preceq_{\tilde{f}}^{\text{part}})$ under the continuous map $[\mathbf{R}] \mapsto ([\mathbf{R}], [\mathbf{P}])$. Order $\mathcal{C}$ by reverse inclusion. Let $\mathcal{K} \subset \mathcal{C}$ be a chain for this order, *i.e.* a nested family of nonempty closed sets. By compactness, $\bigcap_{C \in \mathcal{K}} C \neq \emptyset$; Pick $[\mathbf{P}] \in \bigcap_{C \in \mathcal{K}} C$. Fix $C \in \mathcal{K}$. Since $C = \downarrow[\mathbf{Q}_0]$ for some $[\mathbf{Q}_0] \in \mathbf{Q}_{\tilde{f}}$ and $[\mathbf{P}] \in C$, we have $[\mathbf{P}] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{Q}_0]$. If $[\mathbf{R}] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}]$, then by transitivity $[\mathbf{R}] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{Q}_0]$, hence $[\mathbf{R}] \in C$. Thus $\downarrow[\mathbf{P}] \subseteq C$. Hence $\downarrow[\mathbf{P}]$ is an upper bound of $\mathcal{K}$ in the reverse-inclusion order. Therefore, by Zorn's lemma, $\mathcal{C}$ has a maximal element $\downarrow[\mathbf{P}_*]$ (with respect to reverse inclusion). We claim that $[\mathbf{P}_*]$ is $\preceq_{\tilde{f}}^{\text{part}}$ -minimal in $\mathbf{Q}_{\tilde{f}}$. Indeed, if $[\mathbf{R}] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}_*]$, then $\downarrow[\mathbf{R}] \subseteq \downarrow[\mathbf{P}_*]$. By maximality of $\downarrow[\mathbf{P}_*]$ in $\mathcal{C}$ (reverse inclusion), we must have $\downarrow[\mathbf{R}] = \downarrow[\mathbf{P}_*]$, which implies $[\mathbf{P}_*] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{R}]$. By antisymmetry of $\preceq_{\tilde{f}}^{\text{part}}$ on $\mathbf{Q}_{\tilde{f}}$, $[\mathbf{R}] = [\mathbf{P}_*]$. Hence $[\mathbf{P}_*]$ is minimal. The existence of a maximal element follows analogously by applying the same argument to the closed principal upper sets $\uparrow[\mathbf{P}] := \{[\mathbf{R}] \in \mathbf{Q}_{\tilde{f}} : [\mathbf{P}] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{R}]\}$. Finally, assume $\mathbf{Q}_{\tilde{f}}$ is down-directed. Let $[\mathbf{P}_*]$ be a $\preceq_{\tilde{f}}^{\text{part}}$ -minimal element. For any $[\mathbf{P}] \in \mathbf{Q}_{\tilde{f}}$, by down-directedness there exists $[\mathbf{R}] \in \mathbf{Q}_{\tilde{f}}$ with $[\mathbf{R}] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}_*]$ and $[\mathbf{R}] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}]$. By minimality of $[\mathbf{P}_*]$ we must have $[\mathbf{R}] = [\mathbf{P}_*]$, hence $[\mathbf{P}_*] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}]$. Thus $[\mathbf{P}_*]$ is the least element. The argument for a greatest element is analogous using up-directedness. $\square$

Let us add a discussion on why the conditions in Lemma 2.9 ensure that the greatest and least elements for $\mathbf{Q}_{\tilde{f}}$ exist. Compactness with respect to the topology $\tau_{\preceq_{\tilde{f}}^{\text{part}}}$ can be read as a “no escape” condition: Whenever we approximate an extremal element by moving along a chain, the corresponding limit points – when they exist – remain inside $\mathbf{Q}_{\tilde{f}}$. In other words, $\tau_{\preceq_{\tilde{f}}^{\text{part}}}$ -compactness ensures that $\mathbf{Q}_{\tilde{f}}$ does not “escape to infinity” in a topological sense: It prevents extremal candidates from slipping out of the model class.

Finite $\preceq_{\tilde{f}}^{\text{part}}$ -width restricts the degree of “spread” in the poset $\mathbf{Q}_{\tilde{f}}$, preventing it from having arbitrarily large levels of incomparability. In other words, it ensures that the “pathological elements” of $\mathbf{Q}_{\tilde{f}}$ – *i.e.* the incomparable ones – are controllably few. This helps bridge the gap with the totally ordered case $\preceq_{\tilde{f}}^{\text{tot}}$, where no incomparable elements exist and extremal elements are conceptually straightforward.

The closed-graph condition for $\preceq_{\tilde{f}}^{\text{part}}$ (on the quotient poset $\mathbf{Q}_{\tilde{f}}$) plays a complementary topological role. It says that comparability is stable under limits: If $[\mathbf{P}_n] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}'_n]$ for all n and $[\mathbf{P}_n] \rightarrow [\mathbf{P}]$, $[\mathbf{P}'_n] \rightarrow [\mathbf{P}']$ in $\tau_{\preceq_{\tilde{f}}^{\text{part}}}$, then necessarily $[\mathbf{P}] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}']$. In particular, each principal lower set $\downarrow[\mathbf{P}] := \{[\mathbf{R}] \in \mathbf{Q}_{\tilde{f}} : [\mathbf{R}] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{P}]\}$ and principal upper set $\uparrow[\mathbf{P}] := \{[\mathbf{R}] \in \mathbf{Q}_{\tilde{f}} : [\mathbf{P}] \preceq_{\tilde{f}}^{\text{part}} [\mathbf{R}]\}$ is $\tau_{\preceq_{\tilde{f}}^{\text{part}}}$ -closed. This prevents the order relation from “breaking” at the boundary of $\mathbf{Q}_{\tilde{f}}$ and is what allows compactness arguments (*via* intersections of closed sets) to produce genuine extremal elements *inside* $\mathbf{Q}_{\tilde{f}}$.

Finally, the up-directed and down-directed assumptions ensure that these extremal elements are not merely maximal/minimal in the partial order, but are in fact greatest/least. Down-directedness means that any two models admit a common lower bound in $\mathbf{Q}_{\tilde{f}}$; Thus, once a $\preceq_{\tilde{f}}^{\text{part}}$ -minimal element $[\mathbf{P}]_{\tilde{f}}$ exists, it must lie below every other element (otherwise one could find a strict lower bound for it), hence it is the *least* element. The analogous statement holds for the greatest element under up-directedness. In short, compactness and closed graph yield existence of minimal/maximal elements, while two-sided directedness upgrades them to least/greatest, which is precisely what we need to identify the extremal elements within $\mathbf{Q}_{\tilde{f}}$.

We now give a simple example of a collection $\mathbf{Q}$ and a test function $\tilde{f}$ that satisfy (2.1) and the hypotheses of Lemma 2.9, without the additional complications arising from equivalence classes.

Example 2.10 (A Compact Family with Closed Order and Least/Greatest Elements). Let $E = \{0, 1\}$ with the discrete σ -algebra $\mathcal{F} = 2^E$, and fix the test function $\tilde{f} : E \rightarrow \mathbb{R}$ given by $\tilde{f}(x) = x$. For each parameter $a \in [0, 1]$,

let μ_a be the Bernoulli(a) law on E , *i.e.*

$$\mu_a(\{1\}) = a, \quad \mu_a(\{0\}) = 1 - a.$$

Define the Markov operator K_a by

$$(K_a g)(x) := \int_E g(y) \mu_a(dy), \quad g \in B(E),$$

so that $K_a g$ is constant in x . Notice that K_a is idempotent, $K_a^2 = K_a$.

For $t \geq 0$, define

$$P_t^a := e^{-t}I + (1 - e^{-t})K_a.$$

Then $\mathbf{P}^a = (P_t^a)_{t \geq 0}$ is a Markov semigroup. Indeed, using $K_a^2 = K_a$ and $IK_a = K_aI = K_a$,

$$P_t^a P_s^a = (e^{-t}I + (1 - e^{-t})K_a)(e^{-s}I + (1 - e^{-s})K_a) = e^{-(t+s)}I + (1 - e^{-(t+s)})K_a = P_{t+s}^a.$$

Let

$$\mathbf{Q} := \{\mathbf{P}^a : a \in [0, 1]\}.$$

With respect to the order $\preceq_{\tilde{f}}^{\text{part}}$ (*i.e.* $\mathbf{P} \preceq_{\tilde{f}}^{\text{part}} \mathbf{P}' \iff P_t \tilde{f} \leq P'_t \tilde{f}$ for all $t \geq 0$), we have

$$P_t^a \tilde{f}(x) = e^{-t} \tilde{f}(x) + (1 - e^{-t})\mu_a(\tilde{f}) = e^{-t}x + (1 - e^{-t})a,$$

and therefore

$$\mathbf{P}^a \preceq_{\tilde{f}}^{\text{part}} \mathbf{P}^b \iff a \leq b.$$

Hence $\mathbf{Q}$ is a chain (so it has finite width 1).

Endow $\mathbf{Q}$ with the topology induced by the identification $a \mapsto \mathbf{P}^a$. Then $\mathbf{Q}$ is compact and Hausdorff (homeomorphic to $[0, 1]$), and the order graph is closed since it corresponds to

$$\{(a, b) \in [0, 1]^2 : a \leq b\},$$

which is closed in $[0, 1]^2$. In particular, every order interval $\{\mathbf{P} \in \mathbf{Q} : \mathbf{P}^{a'} \preceq_{\tilde{f}}^{\text{part}} \mathbf{P} \preceq_{\tilde{f}}^{\text{part}} \mathbf{P}^{a''}\}$ is closed.

Finally, $\mathbf{Q}$ admits a least and a greatest element:

$$\underline{\mathbf{P}}_{\tilde{f}} = \mathbf{P}^0, \quad \overline{\mathbf{P}}_{\tilde{f}} = \mathbf{P}^1.$$

We are now ready for the full definition of a compact Imprecise Markov Semigroup.

Definition 2.11 (Compact Imprecise Markov Semigroup). Let $\mathbf{Q}$ be an IMSG, that is, a subset of the space $(B(E)^{\mathbb{R}_+})^{B(E)}$ that satisfies (2.1). Fix any $\tilde{f} \in B(E)$, and recall the quotient poset $\mathbf{Q}_{\tilde{f}} = \mathbf{Q} / \sim_{\tilde{f}}$ introduced above. If the induced order $\preceq_{\tilde{f}}^{\text{tot}}$ on $\mathbf{Q}_{\tilde{f}}$ is total, then we say that $\mathbf{Q}$ is a *compact Imprecise Markov Semigroup* if $\mathbf{Q}_{\tilde{f}}$ is $\tau_{\preceq_{\tilde{f}}^{\text{tot}}}$ -compact.

If instead the induced order $\preceq_{\tilde{f}}^{\text{part}}$ on $\mathbf{Q}_{\tilde{f}}$ is only partial, then we say that $\mathbf{Q}$ is a *compact Imprecise Markov Semigroup* if $\mathbf{Q}_{\tilde{f}}$ has finite $\preceq_{\tilde{f}}^{\text{part}}$ -width, it is $\tau_{\preceq_{\tilde{f}}^{\text{part}}}$ -compact, it is both up- and down-directed, and $\preceq_{\tilde{f}}^{\text{part}}$ has closed graph.

In both cases, we denote by $[\underline{\mathbf{P}}]_{\tilde{f}}$ and $[\overline{\mathbf{P}}]_{\tilde{f}}$ the $\preceq_{\tilde{f}}$ -least and $\preceq_{\tilde{f}}$ -greatest elements of $\mathbf{Q}_{\tilde{f}}$, respectively.

Remark 2.12 (Notation: Quotient Poset, Representatives, and Shorthands). Fix $\tilde{f} \in B(E)$. Whenever we invoke order-theoretic or topological notions associated with $\preceq_{\tilde{f}}$, we work on the quotient poset $\mathbf{Q}_{\tilde{f}} = \mathbf{Q} / \sim_{\tilde{f}}$ endowed with the induced antisymmetric order, and we use bracket notation to emphasise equivalence classes. In particular, $[\underline{\mathbf{P}}]_{\tilde{f}}$ and $[\overline{\mathbf{P}}]_{\tilde{f}}$ denote the $\preceq_{\tilde{f}}$ -least and $\preceq_{\tilde{f}}$ -greatest *equivalence classes* in $\mathbf{Q}_{\tilde{f}}$.

When we drop the brackets and write $\underline{\mathbf{P}}_{\tilde{f}}$ or $\overline{\mathbf{P}}_{\tilde{f}}$, we mean that we have fixed (once and for all in the relevant argument) a representative Markov semigroup in the corresponding class, *i.e.* $\underline{\mathbf{P}}_{\tilde{f}} \in [\underline{\mathbf{P}}]_{\tilde{f}}$ and $\overline{\mathbf{P}}_{\tilde{f}} \in [\overline{\mathbf{P}}]_{\tilde{f}}$. Any operator-level statement involving $\underline{\mathbf{P}}_{\tilde{f}}$ or $\overline{\mathbf{P}}_{\tilde{f}}$ (*e.g.* generators, invariant measures, curvature conditions) is to be understood as a statement about the chosen representatives. For notational simplicity, in the remainder of this section, and in Sections 3 and 4, we adopt the following abuse of notation, $\underline{\mathbf{P}}_{\tilde{f}} \equiv \underline{\mathbf{P}}$ and $\overline{\mathbf{P}}_{\tilde{f}} \equiv \overline{\mathbf{P}}$, with the understanding that the dependence on the chosen function $\tilde{f}$ is implicit.

At this point, the reader may ask themselves why we went to such great lengths to ensure that $\mathbf{Q}$ possesses the least and greatest elements $\underline{\mathbf{P}}$ and $\overline{\mathbf{P}}$, respectively. Informally, the reason is that studying their limiting behavior (that is, studying the limits $\lim_{t \rightarrow \infty} P_t \tilde{f}$ and $\lim_{t \rightarrow \infty} \overline{P}_t \tilde{f}$) will inform us on the limiting behavior of all the elements of $\mathbf{Q}$ evaluated at $\tilde{f}$. This intuition is made formal in Corollaries 3.3 and 4.3. In addition, since $\underline{\mathbf{P}}$ and $\overline{\mathbf{P}}$ are (precise) Markov semigroups, we can use techniques developed in the Markov diffusion operators literature [32] to carry out such a study.

Notice also that the framework of Definition 2.11 is deliberately *non-switching*. That is, we model ambiguity by a *set of precise Markov semigroups* $\mathbf{Q}$, but we do not allow the decision-maker to concatenate two different models over successive time intervals. Concretely, given $\mathbf{P}, \mathbf{P}' \in \mathbf{Q}$, we do *not* assume that there exists $\mathbf{P}'' \in \mathbf{Q}$ such that

$$P''_{t+s} = P'_t \circ P_s, \quad \text{for some } t, s \geq 0,$$

nor do we interpret P'_t as acting on the random output $P_s \tilde{f}$ produced under a different model $\mathbf{P}$. Allowing such concatenations corresponds to admitting *switching* controls, and typically leads to a time-consistent upper (or lower) expectation that is no longer representable by a single precise semigroup: The resulting one-step kernels need not be dominated by a fixed probability kernel and may instead be described by nonlinear transition operators – *e.g.* capacities in the sense of [20].

In Sections 3 and 4 we therefore restrict attention to ambiguity represented by a collection of *precise* semigroups without switching. By contrast, in Section 5 we explicitly *allow* switching and construct the associated (Nisio) nonlinear semigroups, which lets us dispense with compactness while retaining robust long-term behavior.

Furthermore, it is worth mentioning that in Sections 3 and 4 we focus our attention on orders (total and partial) associated with a particular function $\tilde{f}$. Indeed, if one were to consider the “global” comparison

$$\mathbf{P} \preceq \mathbf{P}' \iff P_t f \leq P'_t f, \text{ for all } t \geq 0 \text{ and all } f \in B(E),$$

then this relation is in fact *trivial*: Applying it to $-f$ yields $P_t f \geq P'_t f$, hence $P_t f = P'_t f$ for all $t \geq 0$ and $f \in B(E)$. In other words, such an order collapses to equality and cannot distinguish distinct semigroups. Therefore, to obtain a nontrivial comparison and meaningful extremal elements, one must restrict the function class, which we do here by working with orders induced by a fixed $\tilde{f}$. We note that even under a restricted order, there are families $\mathbf{Q}$ satisfying (2.1) that still fail to possess a least element $\underline{\mathbf{P}}$ with respect to that order, see [9], Example 6.2.

As we can see, the choice of $\tilde{f}$ is then crucial, and it should be driven by the application of interest. For example, as we discussed in Section 1, in the case of a convolutional autoencoder, we can pick $\tilde{f}(z) = e_1^\top \phi_{\text{dec}}(z)$ to study how the probability of the first label evolves over time. In quantitative finance, we could pick $\tilde{f}(z) = \min\{(K - z_i)_+, M\}$, $z \in \mathbb{R}_+^d$, where $(\cdot)_+ = \max(\cdot, 0)$, K is the strike, and M a truncation level. This is a truncated put-option payoff on the i -th asset; The resulting lower and upper semigroups yield worst-case option prices under model uncertainty. In reliability/survival analysis, a viable choice is $\tilde{f}(z) = \text{Ind}_{\{z \leq \theta\}}$, for a failure threshold θ , where $\text{Ind}_{\{\cdot\}}$ denotes the indicator function. The functional tracks the probability that the degradation index X_t remains in the “healthy” region, yielding robust upper and lower survival curves.

Let us now link the least element $\underline{\mathbf{P}}$ with lower previsions [2], sometimes also referred to as lower expectations, a crucial concept – perhaps the most important one – in imprecise probability theory.³ Given a generic set $\mathcal{P}$ of countably additive probability measures on E , the lower prevision for a (measurable bounded) functional $\tilde{f}$ on E is given by $\mathbb{E}(\tilde{f}) = \inf_{P \in \mathcal{P}} \mathbb{E}_P(\tilde{f})$, whenever the expectations exist. Alternatively, a superadditive Choquet capacity $\underline{P}$ can be considered, and the lower prevision is computed according to the Choquet integral $\int_E \tilde{f} d\underline{P}$ [20]. These two definitions only coincide if $\underline{P}$ is 2-monotone [24], Section 2.1.(ii). Conditional lower previsions are defined analogously [2, 36]. In our imprecise Markov semigroup framework of Sections 3 and 4, we have that for every $t \geq 0$, the lower conditional prevision $\mathbb{E}(\tilde{f}(X_t) \mid X_0)$ is equal to $\underline{P}_t \tilde{f}$, which in turn is a linear conditional expectation. This is because $\underline{\mathbf{P}}$ is the least element of $\mathbf{Q}$, so $\underline{\mathbf{P}} \in \mathbf{Q}$, which by (2.1) implies that $\underline{\mathbf{P}}$ is a (precise) Markov semigroup associated with a (linear) Markov process. Working with compact Imprecise Markov Semigroups (IMSGs) lets us represent, within a single object, the agent’s ambiguity about both (i) the long-run law (invariant measure) and (ii) the short-run dynamics (transition probabilities) of a Markov process. In Section 5, instead, we do not require $\mathbf{Q}$ to be compact, so we are able to explicitly relate imprecise Markov semigroups to both the lower envelope $\mathbb{E}(\tilde{f})$ and the Choquet integral $\int_E \tilde{f} d\underline{P}$.

In the imprecise probability literature, ambiguity about invariant behavior is often encoded directly at the level of stationary objects, for instance through an invariant Choquet capacity [6]. Ambiguity about the dynamics is instead typically modeled by sets of transition probabilities (in discrete time) or sets of transition rate matrices (in continuous time), frequently assumed convex so that worst/best-case evaluations can be computed *via* linear-programming-type procedures [9, 10].

A Markov semigroup does not, by itself, specify an initial distribution: It encodes only the transition mechanism $(P_t)_{t \geq 0}$, while an initial law $\nu \in \Delta_E$ must be supplied separately to determine the full process law; Here Δ_E denotes the space of countably additive probability measures on E . In this paper, we represent ambiguity about the *dynamics* by an imprecise Markov semigroup $\mathbf{Q}$, *i.e.* a family of precise semigroups $\mathbf{P} = (P_t)_{t \geq 0}$. Each $\mathbf{P} \in \mathbf{Q}$ determines a candidate collection of transition probabilities and (when it exists) an associated invariant measure $\mu \equiv \mu_{\mathbf{P}}$. Hence $\mathbf{Q}$ captures ambiguity both about transition probabilities and about invariant/long-run behavior, by allowing these objects to vary across $\mathbf{P} \in \mathbf{Q}$. If one also wishes to model ambiguity about the initial distribution, this can be done exogenously by specifying a set of priors $\mathbf{N} \subseteq \Delta_E$ and considering $\{\nu P_t : \nu \in \mathbf{N}, \mathbf{P} \in \mathbf{Q}\}$; Our results primarily focus on the robust long-term behavior induced by uncertainty in the semigroup itself.

As we shall see in Sections 3, 4, and 5, taking an (imprecise Markov) semigroup route allows us to explicitly account for the geometry of the state space E when we study the ergodic behavior of $\overline{\mathbf{P}}$ and $\underline{\mathbf{P}}$ (a perspective that is new to the literature), and to work in continuous time (whereas the majority of the results in the literature pertain the discrete time case). The price we pay in Sections 3 and 4 for these gains is the need for compactness of the equivalence class space under our chosen preorder. This is because, as a consequence of Lemma 2.9, $\overline{\mathbf{P}}, \underline{\mathbf{P}} \in \mathbf{Q}$, and so we can use techniques developed for (precise) Markov semigroups to study the ergodic behavior of $\overline{\mathbf{P}}$ and $\underline{\mathbf{P}}$, and in turn of all the elements of $\mathbf{Q}$. In Section 5, instead, we allow $\mathbf{Q}$ to be non-compact, thus making our approach closer to the frameworks introduced *e.g.* in [18, 19, 37–41],⁴ where nonlinear Markov semigroups (and in particular Markov semigroups associated with nonlinear Markov processes) are studied.

³A similar argument holds for $\overline{\mathbf{P}}$ and upper previsions.

⁴In those papers, the curvature of the state space E is not considered by the authors when deriving their results.

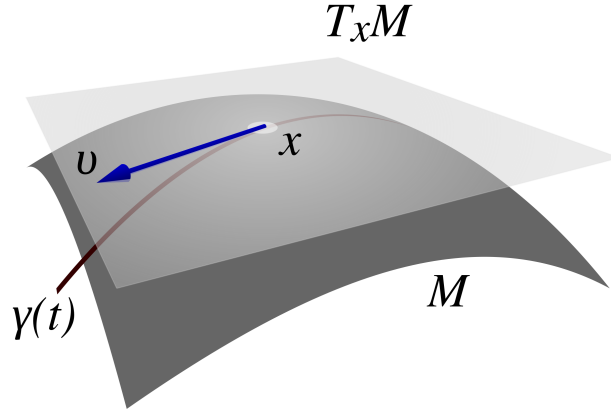


FIGURE 3. In this figure, M is a Riemannian manifold in $\mathbb{R}^3$, $x \in M$ is an element of the manifold, and $T_x M$ is the tangent space of M at x (the set of all tangent vectors to all smooth curves on M passing through x). $\gamma(t)$ is a C^1 curve passing through x , and $v \equiv v(x) = (v^i(x))_{i \in \{1,2,3\}}$ is a vector field.

3. ERGODICITY IN THE RIEMANNIAN MANIFOLD CASE – COMPACT $\mathbf{Q}$

In this section, we begin our inspection on how the geometry of the state space E informs us around the ergodicity of an imprecise Markov semigroup $\mathbf{Q}$. Our journey starts by letting $E = \mathbb{R}^n$, or $E = M$, a complete (with respect to the Riemannian distance) Riemannian manifold [42].⁵

Loosely, a complete Riemannian manifold M can be thought of as a “smooth” subset of $\mathbb{R}^n$, equipped with a complete metric that depends in some sense on the elements $x \in M$. A visual representation of a Riemannian manifold in $\mathbb{R}^3$ is given in Figure 3. In this context, the carré du champ and the iterated carré du champ operators associated with a Markov semigroup are easily computed (see Appendix C.1.1). Recall from Remark 2.12 that, in the following theorem, we have $\underline{\mathbf{P}} \equiv \underline{\mathbf{P}}_{\tilde{f}}$, a fixed chosen representative from $[\underline{\mathbf{P}}]_{\tilde{f}}$, and similarly for $\overline{\mathbf{P}}$.

Theorem 3.1 (Limiting Behavior of $\underline{\mathbf{P}}$ and $\overline{\mathbf{P}}$, E Euclidean space or Riemannian manifold). *Let E be $\mathbb{R}^n$ or a smooth complete connected manifold. Let $\mathcal{A}_0$ be the class of smooth (C^∞) compactly supported functionals f on E .⁶ Fix any $\tilde{f} \in B(E)$, and assume that we can define the partial order $\preceq_{\tilde{f}}^{\text{part}}$ or the total order $\preceq_{\tilde{f}}^{\text{tot}}$ that we introduced in Section 2. Let $\mathbf{Q}$ be a compact imprecise Markov semigroup. Denote by $\overline{\mathbf{P}}$ and $\underline{\mathbf{P}}$ the greatest and least elements of $\mathbf{Q}$,⁷ respectively, with invariant measures $\overline{\mu}$ and $\underline{\mu}$, and associated infinitesimal generators $\overline{L}$ and $\underline{L}$, respectively. Suppose that the following three conditions hold,*

- (i) $\underline{L}$ and $\overline{L}$ are elliptic diffusion operators, symmetric with respect to $\underline{\mu}$ and $\overline{\mu}$,⁸ respectively;
- (ii) $\underline{L}$ and $\overline{L}$ satisfy the Bakry-Émery curvature conditions $CD(\rho_{\underline{L}}, \infty)$ and $CD(\rho_{\overline{L}}, \infty)$, respectively, for some $\rho_{\underline{L}}, \rho_{\overline{L}} \in \mathbb{R}$;
- (iii) The carré du champ algebra for both the carré du champ operators $\underline{\Gamma}$ and $\overline{\Gamma}$ of the Markov generators $\underline{L}$ and $\overline{L}$, respectively, is $\mathcal{A}_0$.

⁵For a primer on Riemannian geometry, we refer the reader to [32], Appendix C.3

⁶Recall that a smooth function on a closed and bounded set (and hence, in this framework, compact by Heine-Borel, or Hopf-Rinow for Riemannian manifold) is itself bounded.

⁷With respect to $\preceq_{\tilde{f}}^{\text{part}}$ or $\preceq_{\tilde{f}}^{\text{tot}}$.

⁸The symmetric assumption tell us that the law of the Markov processes associated with $\underline{\mathbf{P}}$ and $\overline{\mathbf{P}}$ are “reversible in time”.

Then, we have that

$$\forall f \in \mathbb{L}^2(\underline{\mu}), \quad \lim_{t \rightarrow \infty} \underline{P}_t f = \mathbb{E}_{\underline{\mu}}[f] = \int_E f d\underline{\mu}, \quad \underline{\mu}\text{-a.e.}$$

and

$$\forall f \in \mathbb{L}^2(\bar{\mu}), \quad \lim_{t \rightarrow \infty} \bar{P}_t f = \mathbb{E}_{\bar{\mu}}[f] = \int_E f d\bar{\mu}, \quad \bar{\mu}\text{-a.e.}$$

in $\mathbb{L}^2(\underline{\mu})$ and in $\mathbb{L}^2(\bar{\mu})$, respectively. In addition, if the invariant measure for $\underline{\mathbf{P}}$ is equal to that of $\bar{\mathbf{P}}$, and if we denote it by μ , then

$$\forall f \in \mathbb{L}^2(\mu), \quad \lim_{t \rightarrow \infty} \bar{P}_t f = \lim_{t \rightarrow \infty} \underline{P}_t f = \mathbb{E}_{\mu}[f] = \int_E f d\mu, \quad \mu\text{-a.e.} \quad (3.1)$$

in $\mathbb{L}^2(\mu)$.

If E is a complete Riemannian manifold, the Bakry-Émery curvature condition $CD(\rho_{\underline{L}}, \infty)$ in Theorem 3.1.(ii) is related to the notion of Ricci tensor at every point of E , and in turn to the Ricci curvature of the manifold E (for more, we refer the reader to Appendix C.1.1).⁹ We see, then, how the geometry of the state space E influences the ergodic behavior of an imprecise Markov semigroup.

Proof of Theorem 3.1. We begin by pointing out that, as a consequence of Lemma 2.9, $\underline{\mathbf{P}}$ and $\bar{\mathbf{P}}$ are (precise) Markov semigroups, and so the infinitesimal generators $\underline{L}$ and $\bar{L}$ are well-defined.

We first focus on the greatest element $\bar{\mathbf{P}}$ of $\mathbf{Q}$. Recall that the variance $\text{Var}_{\bar{\mu}}(f)$ of function $f \in \mathbb{L}^2(\bar{\mu})$ is given by

$$\text{Var}_{\bar{\mu}}(f) := \int_E f^2 d\bar{\mu} - \left(\int_E f d\bar{\mu} \right)^2, \quad (3.2)$$

and that the Dirichlet form $\bar{\mathcal{E}}(f)$ at $f \in \mathcal{A}_0$ of the carré du champ operator $\bar{\Gamma}$ of the Markov generator $\bar{L}$ is given by

$$\bar{\mathcal{E}}(f) := \int_E \bar{\Gamma}(f) d\bar{\mu}.$$

Pick any $f \in \mathcal{A}_0$. Given our hypotheses, by [32], Theorem 3.2.3, we have that for every $t \geq 0$, the following gradient bound holds

$$\bar{\Gamma}(\bar{P}_t f) \leq e^{-2t\rho_{\bar{L}}} \bar{P}_t(\bar{\Gamma}(f)). \quad (3.3)$$

Then, by [32], Theorem 4.7.2, equation (3.3) implies that the following local Poincaré inequality holds for all $t \geq 0$,

$$\bar{P}_t(f^2) - (\bar{P}_t f)^2 \leq \frac{1 - e^{-2t\rho_{\bar{L}}}}{\rho_{\bar{L}}} \bar{P}_t(\bar{\Gamma}(f)). \quad (3.4)$$

⁹A similar argument holds for the infinitesimal generator $\bar{L}$ associated with Markov semigroup $\bar{\mathbf{P}}$.

Here the factor $\frac{1-e^{-2t\rho_{\bar{L}}}}{\rho_{\bar{L}}}$ needs to be understood as $2t$ when $\rho_{\bar{L}} = 0$. Following [43], Theorem 2.35, taking the limit as $t \rightarrow \infty$ on both sides of (3.4) yields the following Poincaré inequality

$$\text{Var}_{\bar{\mu}}(f) \leq \frac{1}{\rho_{\bar{L}}} \bar{\mathcal{E}}(f). \quad (3.5)$$

Here it is implicitly assumed that $\rho_{\bar{L}} > 0$; We expand on this in Remark 3.2. In addition, recall that for the Dirichlet form $\bar{\mathcal{E}}(f)$ we have that

$$\bar{\mathcal{E}}(f) = - \int_E f \bar{L} f d\bar{\mu}. \quad (3.6)$$

Combining (3.6) with (3.5), we have that if function f is such that $\bar{L}f = 0$, then $\bar{\mathcal{E}}(f) = 0$. In turn, this implies that $\text{Var}_{\bar{\mu}}(f) = 0$, and so the function f must be constant $\bar{\mu}$ -a.e. Moreover, $\mathcal{A}_0$ is a core for $\bar{L}$, *i.e.* it is dense in $\mathcal{D}(\bar{L})$ for the graph norm $\|f\|_{\mathcal{D}(\bar{L})} = \|f\|_{\mathbb{L}^2(\bar{\mu})} + \|\bar{L}f\|_{\mathbb{L}^2(\bar{\mu})}$.¹⁰ Therefore, equations (3.2)–(3.6) established on $\mathcal{A}_0$ extend to $\mathcal{D}(\bar{L})$ by approximation. This implies that every $f \in \mathcal{D}(\bar{L})$ such that $\bar{L}f = 0$ is constant, and so that $\bar{L}$ is ergodic according to [32], Definition 3.1.11. Then, since $\bar{\mu}$ is finite, by [32], Proposition 3.1.13 we have that

$$\forall f \in \mathbb{L}^2(\bar{\mu}), \quad \lim_{t \rightarrow \infty} \bar{P}_t f = \int_E f d\bar{\mu}, \quad \bar{\mu}\text{-a.e.}$$

in $\mathbb{L}^2(\bar{\mu})$. This concludes the first part of the proof. The second part, that is, the one pertaining the least element $\underline{\mathbf{P}}$ of $\mathbf{Q}$, is analogous. The last statement of the theorem is immediate once we let $\bar{\mu} = \underline{\mu} = \mu$. $\square$

Remark 3.2. Let us add a technical remark to Theorem 3.1. If we do not know whether the (reversible invariant) measures $\bar{\mu}$ and $\underline{\mu}$ are finite, then in the statement of condition (ii) of Theorem 3.1 we must require that $\rho_{\bar{L}}, \rho_{\underline{L}} > 0$ [32], Theorem 3.2.7, so that (3.3) holds.

Theorem 3.1 is about the chosen representatives $\underline{\mathbf{P}} \in [\underline{\mathbf{P}}]_{\tilde{f}}$ and $\bar{\mathbf{P}} \in [\bar{\mathbf{P}}]_{\tilde{f}}$: the invariant measures $\underline{\mu}, \bar{\mu}$ and generators $\underline{L}, \bar{L}$ are associated with these representatives. The proof uses only assumptions (i)–(iii) on the representative generator; Extremality is not used beyond selecting which semigroups are considered. Hence, the same convergence conclusion holds for any representative in $\mathbf{Q}$ (in particular, in the extremal classes) satisfying (i)–(iii). Moreover, if two representatives $\mathbf{P}, \mathbf{P}'$ lie in the same class $[\mathbf{P}] = [\mathbf{P}']$ and both satisfy the conclusion with invariant measures μ, μ' , then they necessarily satisfy $\int \tilde{f} d\mu = \int \tilde{f} d\mu'$, since $P_t \tilde{f} = P'_t \tilde{f}$ for all t .

Examples of semigroups for which the conditions (i)–(iii) of Theorem 3.1 are met are the Ornstein-Uhlenbeck and Laguerre semigroups (whose Markov generators satisfy the Bakry-Émery curvature condition with $\rho > 0$; In particular, $\rho = 1$ for the former, and $\rho = 1/2$ for the latter). In addition, an example of two Markov semigroups having the same invariant measure (as required for (3.1) to hold) are the Brownian motion with quadratic potential, and the Ornstein-Uhlenbeck semigroup with volatility parameter equals to $\sqrt{2}$. Indeed, for both, the invariant measure is a Gaussian with mean 0, and variance the inverse of the rate of mean reversion.

The following is an important corollary that subsumes the ergodic behavior of the elements of $\mathbf{Q}$ when applied to $\tilde{f} \in B(E)$. We denote by $\underline{\mu}, \bar{\mu}$ the invariant measures of the chosen representatives of the extremal classes from Theorem 3.1.

Corollary 3.3 (Ergodicity of Compact Imprecise Markov Semigroup $\mathbf{Q}$, E Euclidean space or Riemannian manifold). *Retain the assumptions of Theorem 3.1, and let $\tilde{f} \in B(E)$ be the same functional that we fixed in*

¹⁰This norm and its properties are recalled in Appendix C.2.

Theorem 3.1. Then, for all $\mathbf{P} = (P_t)_{t \geq 0} \in \mathbf{Q}$, we have that

$$\lim_{t \rightarrow \infty} P_t \tilde{f} \geq \int_E \tilde{f} d\mu, \quad \mu\text{-a.e.} \quad (3.7)$$

in $\mathbb{L}^2(\underline{\mu})$, and

$$\lim_{t \rightarrow \infty} P_t \tilde{f} \leq \int_E \tilde{f} d\bar{\mu}, \quad \bar{\mu}\text{-a.e.} \quad (3.8)$$

in $\mathbb{L}^2(\bar{\mu})$. In addition, if the invariant measure for $\mathbf{P}$ is equal to that of $\bar{\mathbf{P}}$, and if we denote it by μ , then

$$\lim_{t \rightarrow \infty} P_t \tilde{f} = \int_E \tilde{f} d\mu, \quad \mu\text{-a.e.} \quad (3.9)$$

in $\mathbb{L}^2(\mu)$.

Proof. Immediate from Theorem 3.1, and the fact that – as a consequence of Lemma 2.9 and Definition 2.11 – $\bar{\mathbf{P}}$ and $\underline{\mathbf{P}}$ are the greatest and least elements of $\mathbf{Q}$, respectively. $\square$

Let us add a few comments on Corollary 3.3. If μ is an invariant measure for a (precise) semigroup $\mathbf{P} = (P_t)_{t \geq 0}$, then by definition $\mu P_t = \mu$ for all $t \geq 0$; Equivalently, $\int_E P_t f d\mu = \int_E f d\mu$ for every bounded f . This is a *stationarity* statement and holds for every t as soon as the initial law is chosen to be μ . Corollary 3.3, however, is of a different nature: It provides *long-term* bounds for the conditional expectation $P_t \tilde{f}(x) = \mathbb{E}[\tilde{f}(X_t) \mid X_0 = x]$ (uniformly in the initial state, and robustly over $\mathbf{Q}$), for a fixed $\tilde{f}$, which is not implied by invariance alone and typically requires mixing/ergodicity assumptions. In particular, in the case $\underline{\mu} = \bar{\mu} = \mu$, the bounds collapse to the same constant,

$$\lim_{t \rightarrow \infty} \underline{P}_t \tilde{f}(x) = \lim_{t \rightarrow \infty} \bar{P}_t \tilde{f}(x) = \int_E \tilde{f} d\mu,$$

so the imprecision about the long-run expectation of $\tilde{f}$ vanishes, even though the short-run values may differ across models in $\mathbf{Q}$. Finally, we emphasize that the limits in Corollary 3.3 are *not* Cesàro time averages along a single trajectory; Rather, they concern convergence of the semigroup (one-time marginals) as $t \rightarrow \infty$. Motivated by (3.7)–(3.9), we say that $\mathbf{Q}$ is *$\tilde{f}$ -lower ergodic* if it satisfies (3.7), *$\tilde{f}$ -upper ergodic* if it satisfies (3.8), and *$\tilde{f}$ -ergodic* if it satisfies (3.9).

As we pointed out in Section 2, there are two main differences with respect to the existing literatures on imprecise Markov processes and ergodicity for imprecise probabilities. We are the first to explicitly exploit the geometry of the state space E via the Bakry-Émery curvature condition to derive our results. In addition, in Theorem 3.1 and Corollary 3.3 we represent the scholar's ambiguity about the invariant measure and the transition probability via a compact imprecise Markov semigroup $\mathbf{Q}$. We briefly mention that our approach differs slightly from the imprecise ergodic literature [22–29], where ambiguity around measure μ is typically represented via a Choquet approach involving a superadditive Choquet capacity.

We conclude this section with a corollary that further informs us on the geometrical properties of E in the ergodic regime of $\mathbf{Q}$, when E is a complete Riemannian manifold. Let us denote by E_K^n the complete n -dimensional simply connected space of constant sectional curvature $K \in \mathbb{R}$,¹¹ and by $B(p, r)$ the ball of radius r at point p , defined with respect to the Riemannian distance function. Once again, the values that we study are tied to the choice of the representatives $\underline{\mathbf{P}}$ and $\bar{\mathbf{P}}$.

¹¹The Ricci curvature of E_K^n is constant and equal to $(n - 1)K$.

Corollary 3.4 (Bishop-Gromov Inequality). *Let $E \subseteq \mathbb{R}^n$ be a complete Riemannian manifold. Retain the first part of Theorem 3.1, and assume that condition (ii) holds. Then,*

$$\text{Vol}[B(p, r)] \leq \text{Vol}[B(p_K, r)],$$

for all $p \in E$, all $p_K \in E_K^n$, all $r \in (0, \infty)$, and some K that depends on ρ_L and $\rho_{\bar{L}}$. Here, $\text{Vol}[\cdot]$ denotes the volume operator.

Proof. Given our assumptions, we know that $CD(\rho_L, \infty)$ and $CD(\rho_{\bar{L}}, \infty)$ both hold. Let us define $\rho_\star := \min\{\rho_L, \rho_{\bar{L}}\}$, and put $\rho_\star = (n-1)K$, so that $K = \rho_\star/(n-1)$. The Bakry-Émery curvature assumptions imply that the Ricci tensor at every point of E is bounded from below by $\rho_\star$.¹² In turn, we have that the Ricci curvature Ric of E is lower bounded by $(n-1)K$, that is, $\text{Ric} \geq (n-1)K$. The result then follows from [44]. $\square$

Corollary 3.4 tells us that – under the Bakry-Émery curvature condition (ii) of Theorem 3.1 – the volume of a ball around any element of the manifold E is upper bounded by the volume of a ball around any element of E_K^n having the same radius. Further comparison results like Corollary 3.4 that hold in the ergodic regime of $\mathbf{Q}$ will be the subject of future study.

4. ERGODICITY IN THE GENERAL CASE – COMPACT $\mathbf{Q}$

In this section, our goal is to extend Theorem 3.1 and Corollary 3.3 to the case where $(E, \mathcal{F})$ is an arbitrary good measurable space, while retaining the assumptions that $\mathbf{Q}$ is compact and $\tilde{f}$ is fixed. To do so, we need to introduce the concept of an Imprecise Diffusion Markov Tuple, the imprecise version of a diffusion Markov triple; The latter is recalled in Appendix C.2. Similarly to Section 3, in the following definition, theorem, and corollary we have that $\underline{\mathbf{P}} \equiv \underline{\mathbf{P}}_{\tilde{f}}$, a fixed chosen representative from $[\underline{\mathbf{P}}]_{\tilde{f}}$, and similarly for $\bar{\mathbf{P}}$.

Definition 4.1 (Imprecise Diffusion Markov (5-)Tuple). Fix any $\tilde{f} \in B(E)$, and assume that we can define the partial order $\preceq_{\tilde{f}}^{\text{part}}$ or the total order $\preceq_{\tilde{f}}^{\text{tot}}$ that we introduced in Section 2. An *imprecise diffusion Markov tuple* $(E, \bar{\mu}, \underline{\mu}, \bar{\Gamma}, \underline{\Gamma})$ is a 5-tuple such that the following holds.

- $(E, \mathcal{F})$ is a good measurable space.
- $\bar{\mu}$ and $\underline{\mu}$ are $(\sigma$ -finite) measures on $\mathcal{F}$.
- $\mathcal{A}_0^{\bar{\mu}}$ and $\mathcal{A}_0^{\underline{\mu}}$ are vector spaces of bounded measurable functionals dense in all $\mathbb{L}^p(\bar{\mu})$ and all $\mathbb{L}^p(\underline{\mu})$ spaces, $1 \leq p < \infty$, respectively, and stable under products.
- $(E, \bar{\mu}, \bar{\Gamma})$ and $(E, \underline{\mu}, \underline{\Gamma})$ are diffusion Markov triples having associated algebras $\mathcal{A}_0^{\bar{\mu}}$ and $\mathcal{A}_0^{\underline{\mu}}$, respectively, and carré du champs operators $\bar{\Gamma}$ and $\underline{\Gamma}$, respectively.
- The Markov semigroups $\bar{\mathbf{P}} = (\bar{P}_t)_{t \geq 0}$ and $\underline{\mathbf{P}} = (\underline{P}_t)_{t \geq 0}$ associated with $\bar{\Gamma}$ and $\underline{\Gamma}$ are the greatest and least elements of a compact imprecise Markov semigroup $\mathbf{Q}$,¹³ respectively.

We are now ready for the main result of this section. It hinges on the concepts of adjoint operator, self-adjointness, essential self-adjointness (ESA), extended algebra $\mathcal{A}$ of the algebra $\mathcal{A}_0$ associated with a diffusion Markov triple, connexity, and weak hypo-ellipticity. They are brushed off in Appendix C.2. As its proof technique is similar to that of Theorem 3.1 (despite exploiting different, more general results), we defer it to Appendix A.

Theorem 4.2 (Limiting Behavior of $\underline{\mathbf{P}}$ and $\bar{\mathbf{P}}$, General). *Fix any $\tilde{f} \in B(E)$, and assume that we can define the partial order $\preceq_{\tilde{f}}^{\text{part}}$ or the total order $\preceq_{\tilde{f}}^{\text{tot}}$ that we introduced in Section 2. Let $(E, \bar{\mu}, \underline{\mu}, \bar{\Gamma}, \underline{\Gamma})$ be an imprecise diffusion Markov tuple, and denote by $\mathcal{A}^{\bar{\mu}}$ and $\mathcal{A}^{\underline{\mu}}$ the extended algebras of $\mathcal{A}_0^{\bar{\mu}}$ and $\mathcal{A}_0^{\underline{\mu}}$ associated with $(E, \bar{\mu}, \bar{\Gamma})$ and $(E, \underline{\mu}, \underline{\Gamma})$, respectively. If the connexity and weak hypo-ellipticity assumptions hold for both $(E, \bar{\mu}, \bar{\Gamma})$ and $(E, \underline{\mu}, \underline{\Gamma})$, and if the Bakry-Émery curvature conditions $CD(\rho_{\bar{L}}, \infty)$ and $CD(\rho_L, \infty)$ are met by both the*

¹²This is discussed in Appendix C.1.1.

¹³With respect to $\preceq_{\tilde{f}}^{\text{part}}$ or $\preceq_{\tilde{f}}^{\text{tot}}$.

infinitesimal generators $\bar{L}$ and $\underline{L}$ associated with $\bar{\Gamma}$ and $\underline{\Gamma}$, respectively, for some $\rho_{\bar{L}}, \rho_{\underline{L}} \in \mathbb{R}$, then the following hold

$$\forall f \in \mathbb{L}^2(\underline{\mu}), \quad \lim_{t \rightarrow \infty} \underline{P}_t f = \mathbb{E}_{\underline{\mu}}[f] = \int_E f d\underline{\mu} \quad \text{in } \mathbb{L}^2(\underline{\mu}), \quad \underline{\mu}\text{-a.e.}$$

and

$$\forall f \in \mathbb{L}^2(\bar{\mu}), \quad \lim_{t \rightarrow \infty} \bar{P}_t f = \mathbb{E}_{\bar{\mu}}[f] = \int_E f d\bar{\mu} \quad \text{in } \mathbb{L}^2(\bar{\mu}), \quad \bar{\mu}\text{-a.e.}$$

In addition, if $\bar{\mu} = \underline{\mu} = \mu$, then

$$\forall f \in \mathbb{L}^2(\mu), \quad \lim_{t \rightarrow \infty} \bar{P}_t f = \lim_{t \rightarrow \infty} \underline{P}_t f = \mathbb{E}_{\mu}[f] = \int_E f d\mu \quad \text{in } \mathbb{L}^2(\mu), \quad \mu\text{-a.e.}$$

The connexity and weak hypo-ellipticity assumptions allow to “import” some of the geometric structure of the Riemannian manifolds considered in Section 3 to the more general diffusion Markov triple setting, and to relax the requirement in Theorem 3.1 that $\underline{L}$ and $\bar{L}$ are elliptic operators. This is expanded upon in Appendix C.2.

Similarly to what we pointed out in Remark 3.2, if we do not know whether the (reversible invariant) measures $\bar{\mu}$ and $\underline{\mu}$ are finite, then in the statement of Theorem 4.2 we must require that $\rho_{\bar{L}}, \rho_{\underline{L}} > 0$ [32], Theorem 3.3.23.

Similarly to Corollary 3.3, the following is an important result that subsumes the ergodic behavior of the elements of a compact IMMSG $\mathbf{Q}$ when applied to $\tilde{f} \in B(E)$.

Corollary 4.3 (Ergodicity of Compact Imprecise Markov Semigroup $\mathbf{Q}$, General). *Retain the assumptions of Theorem 4.2, and let $\tilde{f} \in B(E)$ be the same functional that we fixed in Theorem 4.2. Then, for all $\mathbf{P} = (P_t)_{t \geq 0} \in \mathbf{Q}$, we have that*

$$\lim_{t \rightarrow \infty} P_t \tilde{f} \geq \int_E \tilde{f} d\underline{\mu}, \quad \underline{\mu}\text{-a.e.}$$

in $\mathbb{L}^2(\underline{\mu})$, and

$$\lim_{t \rightarrow \infty} P_t \tilde{f} \leq \int_E \tilde{f} d\bar{\mu}, \quad \bar{\mu}\text{-a.e.}$$

in $\mathbb{L}^2(\bar{\mu})$. In addition, if the invariant measure for $\mathbf{P}$ is equal to that of $\bar{\mathbf{P}}$, and if we denote it by μ , then

$$\lim_{t \rightarrow \infty} P_t \tilde{f} = \int_E \tilde{f} d\mu, \quad \mu\text{-a.e.}$$

in $\mathbb{L}^2(\mu)$.

Proof. Immediate from Theorem 4.2, and the fact that – as a consequence of Lemma 2.9 and Definition 2.11 – $\bar{\mathbf{P}}$ and $\underline{\mathbf{P}}$ are the greatest and least elements of the compact IMMSG $\mathbf{Q}$, respectively. $\square$

5. ERGODICITY IN THE GENERAL CASE – $\mathbf{Q}$ NOT NECESSARILY COMPACT

In this section, we show that the long-term conclusions suggested by our imprecise Markov semigroup viewpoint can be obtained *without* requiring the existence of least/greatest semigroups in $\mathbf{Q}$ (hence without any

compactness assumption; We also do without the notational convention of Remark 2.12). To do so, we need the following.

Let (E, d) be a Polish metric space, $\mathcal{F}$ its Borel σ -algebra, and assume

$$\text{diam}(E) := \sup_{x, y \in E} d(x, y) =: D < \infty. \quad (5.1)$$

For instance, (5.1) holds when E is compact under d . To avoid measurability pathologies for pointwise suprema over uncountable index sets (and over switching times), we work with the universal completion $\mathcal{F}^*$ of $\mathcal{F}$, and interpret envelopes as universally measurable; We expand on this in Appendix B. We require the following.

Assumption 5.1 (Standard Borel Parametrization and Jointly Measurable Kernels). The family $\mathbf{Q}$ satisfying (2.2) is parameterized as $\mathbf{Q} = \{\mathbf{P}^\theta\}_{\theta \in \Theta}$ for a standard Borel space Θ , where $\mathbf{P}^\theta = (P_t^\theta)_{t \geq 0}$ is a Markov semigroup with transition kernels $p_t^\theta(x, dy)$. For every $A \in \mathcal{F}$, the map $(t, x, \theta) \mapsto p_t^\theta(x, A)$ is Borel measurable on $\mathbb{R}_+ \times E \times \Theta$.

Assumption 5.1 is a standard measurability hypothesis from stochastic control: It rules out pathological nonmeasurable envelopes when taking the supremum over θ and when switching along partitions. It is typically satisfied for parametrized SDE/semigroup families with measurable dependence on θ , and imposes no compactness requirements on Θ .

Let Δ_E be the set of Borel probability measures on $(E, \mathcal{F})$ and $\Delta_E^1 := \{\mu \in \Delta_E : \int_E d(x_0, x) \mu(dx) < \infty\}$ for some (hence any) $x_0 \in E$. Under (5.1), we have $\Delta_E^1 = \Delta_E$. For a bounded function $f : E \rightarrow \mathbb{R}$ define the oscillation

$$\text{osc}(f) := \sup_{x \in E} f(x) - \inf_{x \in E} f(x).$$

Denote by $\text{Lip}_b(E)$ the space of bounded d -Lipschitz functions, equipped with the seminorm

$$\text{Lip}(f) := \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)} \in [0, \infty).$$

Let us now recall the definition of the 1-Wasserstein metric; It is instrumental to the second assumption we need to study the ergodicity of a non-compact $\mathbf{Q}$.

Definition 5.2 (1-Wasserstein Distance). For $\mu, \nu \in \Delta_E^1$, let $W_1(\mu, \nu)$ denote the 1-Wasserstein distance, characterized by the Kantorovich-Rubinstein duality

$$W_1(\mu, \nu) = \sup \left\{ \int_E \varphi d(\mu - \nu) : \varphi \in \text{Lip}_b(E), \text{Lip}(\varphi) \leq 1 \right\}. \quad (5.2)$$

Equivalently,

$$W_1(\mu, \nu) = \sup \left\{ \left| \int_E \varphi d(\mu - \nu) \right| : \varphi \in \text{Lip}_b(E), \text{Lip}(\varphi) \leq 1 \right\},$$

since φ is admissible if and only if $-\varphi$ is admissible, that is, if φ has Lipschitz constant ≤ 1 , then $-\varphi$ also has Lipschitz constant ≤ 1 , hence both are allowed as test functions in the supremum.

Assumption 5.3 (Uniform Exponential 1-Wasserstein Contraction). There exists $\lambda > 0$ such that for every $\theta \in \Theta$ and every $t \geq 0$,

$$W_1(\mu P_t^\theta, \nu P_t^\theta) \leq e^{-\lambda t} W_1(\mu, \nu), \quad \forall \mu, \nu \in \Delta_E^1, \quad (5.3)$$

where $\mu P_t^\theta(A) := \int_E p_t^\theta(x, A) \mu(dx)$, for all $A \in \mathcal{F}$, and similarly for νP_t^θ .

Assumption 5.3 is a *uniform exponential mixing/stability condition in the 1-Wasserstein distance W_1* : For every candidate model $\mathbf{P}^\theta$, the semigroup contracts distances between laws at rate $e^{-\lambda t}$, uniformly over θ .

In many diffusion settings, a Bakry-Émery curvature lower bound can yield a contraction of the form (5.3) (possibly in W_1 , in the 2-Wasserstein distance W_2 , or for the intrinsic distance associated with the carré du champ), *via* gradient estimates and the duality between gradient bounds and Wasserstein contraction. The following is an example.

Example 5.4 (Ornstein-Uhlenbeck Family: A θ -Parametrized Curvature-Driven W_1 Contraction). Let $E = \mathbb{R}^d$ with Euclidean distance $d(x, y) = \|x - y\|$. Of course, $\mathbb{R}^d$ does not have finite diameter: Assumption (5.1) is not needed for this example. Fix a parameter set Θ and, for each $\theta \in \Theta$, consider the Ornstein-Uhlenbeck generator

$$L^\theta f(x) = \text{tr}(a^\theta \nabla^2 f(x)) - \langle b^\theta x, \nabla f(x) \rangle,$$

where a^θ is a symmetric positive semidefinite $d \times d$ matrix and b^θ is a $d \times d$ matrix whose spectrum is contained in $(0, \infty)$ (so the drift is mean-reverting). Let $\mathbf{P}^\theta = (P_t^\theta)_{t \geq 0}$ be the associated Markov semigroup.

Equivalently, $\mathbf{P}^\theta$ is the transition semigroup of the linear SDE

$$dX_t = -b^\theta X_t dt + \sqrt{2a^\theta} dW_t,$$

where $(W_t)_{t \geq 0}$ is a standard d -dimensional Brownian motion, and the Mehler formula yields, for bounded measurable f ,

$$P_t^\theta f(x) = \mathbb{E} \left[f(e^{-tb^\theta} x + Z_t^\theta) \right], \quad Z_t^\theta \sim \mathcal{N}(0, \Sigma_t^\theta),$$

where $\mathcal{N}(0, \Sigma_t^\theta)$ denotes a Gaussian distribution with mean 0 and covariance $\Sigma_t^\theta = \int_0^t e^{-sb^\theta} (2a^\theta) e^{-s(b^\theta)^\top} ds$.

If $f \in \text{Lip}_b(E)$, then for all $x, y \in \mathbb{R}^d$,

$$|P_t^\theta f(x) - P_t^\theta f(y)| \leq \mathbb{E} \left[\left| f(e^{-tb^\theta} x + Z_t^\theta) - f(e^{-tb^\theta} y + Z_t^\theta) \right| \right] \leq \text{Lip}(f) \|e^{-tb^\theta} (x - y)\|.$$

Hence

$$\text{Lip}(P_t^\theta f) \leq \|e^{-tb^\theta}\|_{\text{op}} \text{Lip}(f) \quad \forall t \geq 0,$$

where $\|\cdot\|_{\text{op}}$ denotes the operator norm. If, in addition, there exists $\lambda > 0$ such that

$$\sup_{\theta \in \Theta} \|e^{-tb^\theta}\|_{\text{op}} \leq e^{-\lambda t} \quad \forall t \geq 0,$$

(e.g. it suffices that $\inf_{\theta \in \Theta} \lambda_{\min}((b^\theta + (b^\theta)^\top)/2) \geq \lambda$), then for all $\theta \in \Theta$ and $t \geq 0$,

$$\text{Lip}(P_t^\theta f) \leq e^{-\lambda t} \text{Lip}(f).$$

By Kantorovich-Rubinstein duality, this is equivalent to the uniform 1-Wasserstein contraction

$$W_1(\mu P_t^\theta, \nu P_t^\theta) \leq e^{-\lambda t} W_1(\mu, \nu), \quad \forall \mu, \nu \in \Delta_E^1, \forall t \geq 0, \forall \theta \in \Theta,$$

i.e. Assumption 5.3 holds with such a λ .

Example 5.4 can be viewed as a prototype of the general principle “Bakry-Émery curvature lower bounds $\implies$ gradient estimates for $(P_t)_{t \geq 0} \implies$ Wasserstein contraction of $\mu \mapsto \mu P_t$ ”, with a precise equivalence between suitable gradient bounds and W_p -contraction given by Kuwada-type duality [45, 46]. We formalize such a principle in Lemma 5.6.

In this section, we take (5.3) as a compactness-free replacement for the existence of extremal semigroups: It is strong enough to yield uniform ergodic control (hence a robust limit) for the sublinear Nisio semigroup introduced in Section 5.2. Moreover, (5.3) is verifiable in concrete model classes *via* metric contractive couplings or, in diffusion settings, *via* gradient estimates (*e.g.* stemming from Bakry-Émery curvature lower bounds) and their dual formulation in Wasserstein distance. We now record the dual (Lipschitz) form of Assumption 5.3.

Lemma 5.5 (Wasserstein Contraction $\iff$ Lipschitz Contraction). *We have that (5.3) holds if and only if*

$$\text{Lip}(P_t^\theta f) \leq e^{-\lambda t} \text{Lip}(f), \quad \forall f \in \text{Lip}_b(E), \forall \theta \in \Theta, \forall t \geq 0. \quad (5.4)$$

Proof. Assume (5.3). Fix θ, t and $f \in \text{Lip}_b(E)$. For $x \neq y$, by (5.2) and (5.3),

$$|P_t^\theta f(x) - P_t^\theta f(y)| = \left| \int_E f d(\delta_x P_t^\theta - \delta_y P_t^\theta) \right| \leq \text{Lip}(f) W_1(\delta_x P_t^\theta, \delta_y P_t^\theta) \leq \text{Lip}(f) e^{-\lambda t} d(x, y).$$

Taking the supremum over $x \neq y$ gives (5.4).

Conversely, assume (5.4). Fix $\mu, \nu \in \Delta_E^1$ and $\varphi \in \text{Lip}_b(E)$ with $\text{Lip}(\varphi) \leq 1$. Then, by (5.4), $\text{Lip}(P_t^\theta \varphi) \leq e^{-\lambda t}$ and

$$\left| \int_E \varphi d(\mu P_t^\theta - \nu P_t^\theta) \right| = \left| \int_E (P_t^\theta \varphi) d(\mu - \nu) \right| \leq \text{Lip}(P_t^\theta \varphi) W_1(\mu, \nu) \leq e^{-\lambda t} W_1(\mu, \nu).$$

Taking the supremum over φ with $\text{Lip}(\varphi) \leq 1$ and using (5.2) yields (5.3). $\square$

The next result uses the equivalence in Lemma 5.5 to show that a (uniform) Bakry-Émery curvature condition on $\mathbf{Q}$ implies our Assumption 5.3.

Lemma 5.6 (Bakry-Émery Curvature $\implies W_1$ -Contraction). *Let (E, μ, Γ) be a diffusion Markov triple with Markov semigroup $(P_t)_{t \geq 0}$ and generator L . Let $\mathcal{A}$ be an algebra of bounded functionals, stable under products and invariant under $(P_t)_{t \geq 0}$, such that Γ is well-defined on $\mathcal{A}$. Define the intrinsic (pseudo-)distance on E associated with Γ by*

$$d_\Gamma(x, y) := \sup \left\{ u(x) - u(y) : u \in \mathcal{A}, \|\sqrt{\Gamma(u)}\|_\infty \leq 1 \right\} \in [0, \infty]. \quad (5.5)$$

Assume that d_Γ is finite and separates points, so that (E, d_Γ) is a metric space, and that (E, d_Γ) is Polish with Borel σ -algebra $\mathcal{F}$. Assume also the following density condition,

$$\forall \varphi \in \text{Lip}_b(E, d_\Gamma) \text{ with } \text{Lip}_{d_\Gamma}(\varphi) \leq 1, \exists (f_n)_{n \geq 1} \subset \left\{ f \in \mathcal{A} : \|\sqrt{\Gamma(f)}\|_\infty \leq 1 \right\} \text{ such that } \|f_n - \varphi\|_\infty \rightarrow 0. \quad (5.6)$$

For ν_1, ν_2 Borel probability measures with finite first moment with respect to d_Γ , let $W_{1, d_\Gamma}(\nu_1, \nu_2)$ denote the 1-Wasserstein distance associated with d_Γ – equivalently given by the Kantorovich-Rubinstein duality on (E, d_Γ) .

If the Bakry-Émery curvature condition $CD(\rho, \infty)$ holds for some $\rho \in \mathbb{R}$, then for all $t \geq 0$ and all such ν_1, ν_2 ,

$$W_{1, d_\Gamma}(\nu_1 P_t, \nu_2 P_t) \leq e^{-\rho t} W_{1, d_\Gamma}(\nu_1, \nu_2). \quad (5.7)$$

In particular, for an imprecise Markov semigroup $\mathbf{Q} = \{\mathbf{P}^\theta = (P_t^\theta)_{t \geq 0}\}_{\theta \in \Theta}$, assume that for each θ we have a diffusion Markov triple $(E, \mu_\theta, \Gamma_\theta)$ with intrinsic distance d_{Γ_θ} , and that

$$d_{\Gamma_\theta} = d_\Gamma \quad \text{for all } \theta \in \Theta \quad \text{and} \quad \inf_{\theta \in \Theta} \rho_\theta \geq \lambda > 0,$$

where $CD(\rho_\theta, \infty)$ holds for the θ -model. Then, Assumption 5.3 holds on (E, d_Γ) with rate λ .

Assumption (5.6) is standard density property: It holds in many standard settings. For example, in the Riemannian manifold case it follows from smooth approximation of Lipschitz functions with almost-preserved Lipschitz constant [47].

Proof of Lemma 5.6. Step 1: Gradient Estimate under $CD(\rho, \infty)$. Assume that the Bakry-Émery curvature condition $CD(\rho, \infty)$ holds. Then the Bakry-Émery gradient estimate holds on $\mathcal{A}$: for all $t \geq 0$ and all $f \in \mathcal{A}$,

$$\Gamma(P_t f) \leq e^{-2\rho t} P_t \Gamma(f) \quad \mu\text{-a.e.} \quad (5.8)$$

Step 2: Contraction of the Γ -Seminorm. If $\|\sqrt{\Gamma(f)}\|_\infty \leq 1$, then $0 \leq \Gamma(f) \leq 1$ μ -a.e., hence by positivity and constant-preservation of P_t ,

$$0 \leq P_t \Gamma(f) \leq 1 \quad \mu\text{-a.e.}$$

Combining this with (5.8) yields

$$\Gamma(P_t f) \leq e^{-2\rho t} \quad \mu\text{-a.e.}, \quad \text{so} \quad \|\sqrt{\Gamma(P_t f)}\|_\infty \leq e^{-\rho t}.$$

Equivalently, if $\|\sqrt{\Gamma(f)}\|_\infty \leq 1$, then

$$\left\| \sqrt{\Gamma(e^{\rho t} P_t f)} \right\|_\infty \leq 1. \quad (5.9)$$

Step 3: Intrinsic Lipschitz Bound. We claim that for every $g \in \mathcal{A}$,

$$\text{Lip}_{d_\Gamma}(g) \leq \|\sqrt{\Gamma(g)}\|_\infty. \quad (5.10)$$

Indeed, if $\|\sqrt{\Gamma(g)}\|_\infty < \infty$, then for any $x, y \in E$ the function $u := g/\|\sqrt{\Gamma(g)}\|_\infty$ satisfies $\|\sqrt{\Gamma(u)}\|_\infty \leq 1$, hence by the definition (5.5) of d_Γ ,

$$\frac{g(x) - g(y)}{\|\sqrt{\Gamma(g)}\|_\infty} = u(x) - u(y) \leq d_\Gamma(x, y).$$

Swapping x, y gives $|g(x) - g(y)| \leq \|\sqrt{\Gamma(g)}\|_\infty d_\Gamma(x, y)$, proving (5.10).

As a consequence of (5.10) and Step 2, for every $f \in \mathcal{A}$ with $\|\sqrt{\Gamma(f)}\|_\infty < \infty$ we have the estimate

$$\text{Lip}_{d_\Gamma}(P_t f) \leq \|\sqrt{\Gamma(P_t f)}\|_\infty \leq e^{-\rho t} \|\sqrt{\Gamma(f)}\|_\infty. \quad (5.11)$$

Step 4: W_{1,d_Γ} -Contraction via Kantorovich-Rubinstein Duality and a Core Density. Let

$$\mathbf{F} := \left\{ f \in \mathcal{A} : \|\sqrt{\Gamma(f)}\|_\infty \leq 1 \right\}.$$

By (5.10), every $f \in \mathbf{F}$ is 1-Lipschitz w.r.t. d_Γ , hence is an admissible test function in the Kantorovich-Rubinstein duality for W_{1,d_Γ} .

Fix ν_1, ν_2 with finite first moment with respect to d_Γ . For any $f \in \mathbf{F}$, define $g_t := e^{\rho t} P_t f$. By (5.9) we have $g_t \in \mathbf{F}$, and therefore

$$\int_E P_t f d(\nu_1 - \nu_2) = e^{-\rho t} \int_E g_t d(\nu_1 - \nu_2) \leq e^{-\rho t} W_{1,d_\Gamma}(\nu_1, \nu_2),$$

where the last inequality follows from (5.2) since $\text{Lip}_{d_\Gamma}(g_t) \leq 1$. Taking the supremum over $f \in \mathbf{F}$ yields

$$\sup_{f \in \mathbf{F}} \int_E P_t f d(\nu_1 - \nu_2) \leq e^{-\rho t} W_{1,d_\Gamma}(\nu_1, \nu_2). \quad (5.12)$$

To pass from the restricted supremum to the full Kantorovich-Rubinstein supremum, we use Assumption (5.6). Let $\varphi \in \text{Lip}_b(E, d_\Gamma)$ with $\text{Lip}_{d_\Gamma}(\varphi) \leq 1$ and pick $f_n \in \mathbf{F}$ with $\|f_n - \varphi\|_\infty \rightarrow 0$. Since P_t is a contraction on L^∞ , we have that $\|P_t f_n - P_t \varphi\|_\infty \leq \|f_n - \varphi\|_\infty \rightarrow 0$. Hence, from (5.12) it follows that

$$\int_E P_t \varphi d(\nu_1 - \nu_2) = \lim_{n \rightarrow \infty} \int_E P_t f_n d(\nu_1 - \nu_2) \leq e^{-\rho t} W_{1,d_\Gamma}(\nu_1, \nu_2).$$

Taking the supremum over all such φ and using Kantorovich-Rubinstein duality (5.2) (with respect to d_Γ) gives

$$W_{1,d_\Gamma}(\nu_1 P_t, \nu_2 P_t) \leq e^{-\rho t} W_{1,d_\Gamma}(\nu_1, \nu_2),$$

which is (5.7).

Final statement for $\mathbf{Q}$. For an imprecise Markov semigroup $\mathbf{Q} = \{\mathbf{P}^\theta = (P_t^\theta)_{t \geq 0}\}_{\theta \in \Theta}$, apply the above argument to each θ . If $d_{\Gamma_\theta} = d_\Gamma$ for all θ and $\inf_{\theta \in \Theta} \rho_\theta \geq \lambda > 0$, then (5.7) holds with rate λ uniformly in θ , which is exactly Assumption 5.3 on (E, d_Γ) . $\square$

The key structural requirement in Lemma 5.6 is the compatibility between the carré du champ and the metric, here encoded by the intrinsic distance d_Γ – or more generally by a domination of $\sqrt{\Gamma(f)}$ in terms of $\text{Lip}(f)$. This is not overly restrictive: It is automatically satisfied in the classical Riemannian setting (where $\Gamma(f) = |\nabla f|^2$ and d_Γ coincides with the geodesic distance) and, more broadly, for strongly local regular Dirichlet forms where d_Γ yields the natural geometry of the diffusion. In such cases, a curvature lower bound $CD(\rho, \infty)$ directly translates into exponential contraction in W_1 for the intrinsic metric.

We also prove that a contraction in the Lipschitz seminorm implies contraction of oscillations.

Lemma 5.7 (Lipschitz Contraction Implies Oscillation Contraction). *Assume (5.1). Then, for every $f \in \text{Lip}_b(E)$,*

$$\text{osc}(f) \leq D \text{Lip}(f).$$

In particular, under Assumption 5.3, for every $\theta \in \Theta$ and $t \geq 0$,

$$\text{osc}(P_t^\theta f) \leq D e^{-\lambda t} \text{Lip}(f), \quad f \in \text{Lip}_b(E).$$

Proof. For any $x, y \in E$, $|f(x) - f(y)| \leq \text{Lip}(f) d(x, y) \leq D \text{Lip}(f)$. Taking $\sup_{x, y}$ gives $\text{osc}(f) \leq D \text{Lip}(f)$. The second claim follows from Lemma 5.5. $\square$

We now turn to the two complementary extensions that avoid compactness of $\mathbf{Q}$. First, in Section 5.1, we analyze *static* upper/lower envelopes (optimizing only at the terminal time), which yield ergodic limits without

any attainment. That is, without needing the supremum/infimum to be achieved by some $\theta^* \in \Theta$, or equivalently any “extremal” semigroup/model.

Second, in Section 5.2 we construct *time-consistent* (Nisio) nonlinear semigroups by allowing switching, thus going beyond Definition 2.11 which explicitly excludes switches, and prove their robust ergodicity under uniform contraction.

In both sections we retain Assumptions 5.1 and 5.3. By Lemma 5.6, the latter can be ensured by asking the elements of $\mathbf{Q}$ to all have positive Bakry-Émery curvature condition. This viewpoint gives the present analysis a geometric flavor, as illustrated for instance by the Ornstein-Uhlenbeck family in Example 5.4.

5.1. Static envelopes: robust limits on $\text{Lip}_b(E)$

For $f \in \text{Lip}_b(E)$ define the static upper and lower envelopes

$$\bar{T}_t f(x) := \sup_{\theta \in \Theta} P_t^\theta f(x), \quad \underline{T}_t f(x) := \inf_{\theta \in \Theta} P_t^\theta f(x), \quad t \geq 0, \quad (5.13)$$

interpreted as $\mathcal{F}^*$ -measurable envelopes – see Appendix B, where the interested reader can find a discussion on the measurability of $x \mapsto \bar{T}_t f(x)$ and $x \mapsto \underline{T}_t f(x)$, and of the switching operators introduced in Definitions 5.11 and 5.16.

In Imprecise Probability terminology, given a set $\mathcal{P}$ of probability measures on $(E, \mathcal{F})$, the *upper* and *lower* expectations of a bounded observable f are defined by

$$\bar{\mathbb{E}}(f) := \sup_{P \in \mathcal{P}} \int_E f dP, \quad \underline{\mathbb{E}}(f) := \inf_{P \in \mathcal{P}} \int_E f dP.$$

The operators in (5.13) can be viewed as a *dynamic* analogue of this construction: For fixed $t \geq 0$ and initial state $x \in E$, each candidate model θ induces the law $\delta_x P_t^\theta$ of X_t under $\mathbf{P}^\theta$, so that $P_t^\theta f(x) = \int_E f d(\delta_x P_t^\theta)$ is the corresponding (precise) expectation. Thus, the envelopes

$$\bar{T}_t f(x) = \sup_{\theta \in \Theta} P_t^\theta f(x) = \sup_{\theta \in \Theta} \int_E f d(\delta_x P_t^\theta), \quad \underline{T}_t f(x) = \inf_{\theta \in \Theta} \int_E f d(\delta_x P_t^\theta),$$

are exactly the upper and lower expectations of f with respect to the time- t set $\mathcal{P}_t(x) := \{\delta_x P_t^\theta : \theta \in \Theta\}$ of one-time marginals.

We first prove a model-wise ergodic estimate that will later be used to identify the upper and lower bounds for the long-term behavior of the envelopes associated with the imprecise Markov semigroup $\mathbf{Q}$.

Proposition 5.8 (Unique Invariant Law and Uniform Ergodic Bound in W_1). *Let Assumptions 5.1 and 5.3 hold, and assume (5.1). Fix $\theta \in \Theta$. Then, there exists a unique invariant probability measure $\mu_\theta \in \Delta_E$ such that for all $t \geq 0$,*

$$W_1(\delta_x P_t^\theta, \mu_\theta) \leq D e^{-\lambda t}, \quad \forall x \in E. \quad (5.14)$$

Consequently, for every $f \in \text{Lip}_b(E)$,

$$\sup_{x \in E} |P_t^\theta f(x) - \mu_\theta(f)| \leq D e^{-\lambda t} \text{Lip}(f), \quad (5.15)$$

where $\mu_\theta(f) := \int_E f d\mu_\theta$.

Proof. Fix θ and $t_0 > 0$. By (5.3), the map $\Phi : \Delta_E^1 \rightarrow \Delta_E^1$, $\Phi(\nu) = \nu P_{t_0}^\theta$, is a strict contraction in W_1 with constant $e^{-\lambda t_0} < 1$. Since (Δ_E^1, W_1) is complete and $\Delta_E^1 = \Delta_E$ under (5.1), Banach's fixed point theorem yields a unique fixed point μ_θ and $\nu P_{nt_0}^\theta \rightarrow \mu_\theta$ in W_1 .

Notice that invariance extends to all $t \geq 0$: for any $s \geq 0$, $\mu_\theta P_s^\theta$ is also a fixed point of Φ (by the semigroup property), hence equals μ_θ by uniqueness.

For (5.14), by (5.3) and invariance,

$$W_1(\delta_x P_t^\theta, \mu_\theta) = W_1(\delta_x P_t^\theta, \mu_\theta P_t^\theta) \leq e^{-\lambda t} W_1(\delta_x, \mu_\theta).$$

Under (5.1), $W_1(\delta_x, \mu_\theta) = \int_E d(x, y) \mu_\theta(dy) \leq D$, yielding (5.14). Finally, (5.15) follows from (5.2),

$$|P_t^\theta f(x) - \mu_\theta(f)| = \left| \int_E f d(\delta_x P_t^\theta - \mu_\theta) \right| \leq \text{Lip}(f) W_1(\delta_x P_t^\theta, \mu_\theta) \leq D e^{-\lambda t} \text{Lip}(f).$$

□

We are now ready to study the ergodicity of $\mathbf{Q}$.

Theorem 5.9 (Static Envelope Limits Under Wasserstein Contraction). *Let Assumptions 5.1 and 5.3 hold, and assume (5.1). Then for every $f \in \text{Lip}_b(E)$,*

$$\lim_{t \rightarrow \infty} \overline{T}_t f(x) = \sup_{\theta \in \Theta} \mu_\theta(f), \quad (5.16)$$

$$\lim_{t \rightarrow \infty} \underline{T}_t f(x) = \inf_{\theta \in \Theta} \mu_\theta(f), \quad (5.17)$$

and the convergence is uniform in x . Quantitatively,

$$\sup_{x \in E} \left| \overline{T}_t f(x) - \sup_{\theta \in \Theta} \mu_\theta(f) \right| \leq D e^{-\lambda t} \text{Lip}(f), \quad \sup_{x \in E} \left| \underline{T}_t f(x) - \inf_{\theta \in \Theta} \mu_\theta(f) \right| \leq D e^{-\lambda t} \text{Lip}(f).$$

Proof. By Proposition 5.8, for every θ and $x \in E$,

$$P_t^\theta f(x) = \mu_\theta(f) + \varepsilon_{\theta,t}(x), \quad |\varepsilon_{\theta,t}(x)| \leq D e^{-\lambda t} \text{Lip}(f).$$

Taking the supremum (resp. infimum) over θ yields the desired bounds and limits. □

Similarly to Corollaries 3.3 and 4.3, the following is an important result that subsumes the ergodic behavior of the elements of an IMMSG $\mathbf{Q}$ that is not assumed to be compact, and that holds for all bounded Lipschitz functions on a finite-diameter Polish metric state space E . It follows immediately from Theorem 5.9.

Corollary 5.10 (Pointwise Ergodic Limit for Each Model). *Assume the hypotheses of Proposition 5.8. Then, for every $f \in \text{Lip}_b(E)$ and every $\theta \in \Theta$,*

$$\lim_{t \rightarrow \infty} P_t^\theta f(x) = \mu_\theta(f) \quad \text{uniformly in } x \in E.$$

In particular, for all $\theta \in \Theta$ and all $x \in E$,

$$\lim_{t \rightarrow \infty} P_t^\theta f(x) \in \left[\inf_{\theta \in \Theta} \mu_\theta(f), \sup_{\theta \in \Theta} \mu_\theta(f) \right].$$

5.2. A time-consistent sublinear semigroup (Nisio) and robust ergodicity

Notice that the static envelopes $\bar{T}_t$ in Section 5.1 need not satisfy the semigroup property. That is, it need not hold that for all $t, s \geq 0$, $\bar{T}_{t+s} = \bar{T}_t \circ \bar{T}_s$.¹⁴ We now build a time-consistent sublinear semigroup by switching along partitions, which satisfies the semigroup property.

Definition 5.11 (Time-Consistent Switching, Nisio Semigroup). For $t \geq 0$ and $f \in \text{Lip}_b(E)$, define the one-step upper envelope $\bar{T}_t f$ as in (5.13). Let Π_t be the set of all finite partitions $\pi = (t_0, \dots, t_n)$ of $[0, t]$ with $0 = t_0 < \dots < t_n = t$. For $\pi = (t_0, \dots, t_n) \in \Pi_t$, define

$$T_t^\pi f := \bar{T}_{t_n - t_{n-1}} \circ \dots \circ \bar{T}_{t_1 - t_0} f,$$

and define the pasted semigroup $(T_t)_{t \geq 0}$ by the envelope

$$T_t f := \sup_{\pi \in \Pi_t} T_t^\pi f, \quad t \geq 0, \quad (5.18)$$

interpreted as a universally measurable envelope.

The static envelopes $\bar{T}_t f(x) = \sup_{\theta \in \Theta} P_t^\theta f(x)$ and $\underline{T}_t f(x) = \inf_{\theta \in \Theta} P_t^\theta f(x)$ should be viewed as the *one-step* upper and lower expectations of the observable $f(X_t)$ under a family of models: Starting from $X_0 = x$, the set of candidate one-time marginals at time t is $\{\delta_x P_t^\theta : \theta \in \Theta\}$, so $\bar{T}_t f(x)$ (resp. $\underline{T}_t f(x)$) is the upper (resp. lower) expectation of f over that set. The Nisio (switching) semigroup $(T_t)_{t \geq 0}$ refines this viewpoint by allowing *time-consistent* recombination of models: Informally, $T_t f(x)$ corresponds to optimizing over *adapted switching rules* that select, over successive time intervals, which semigroup P^θ governs the evolution, yielding a dynamic programming principle $T_{t+s} = T_t \circ T_s$. In this sense, (T_t) plays the role of a *dynamic* upper expectation (and its lower analogue of a dynamic lower expectation), exactly as in robust control.

This dynamic upper/lower expectation perspective is closely related to Choquet capacities: If we encode ambiguity at time t by set functions on events,

$$\bar{P}_t(A | x) := \sup_{\theta \in \Theta} (\delta_x P_t^\theta)(A), \quad \underline{P}_t(A | x) := \inf_{\theta \in \Theta} (\delta_x P_t^\theta)(A), \quad A \in \mathcal{F},$$

then for bounded measurable f the corresponding upper/lower expectations can be expressed *via* the associated (upper/lower) Choquet expectations. In general, however, the switching/Nisio operators T_t and their lower counterparts are best understood as *time-consistent* (sub/super)linear envelopes of conditional expectations (given by iterated one-step suprema/infima along partitions), rather than as integration with respect to a single fixed capacity on path space; Allowing switches typically leads to a nonlinear, dynamically consistent evaluation that subsumes (and may be strictly richer than) the one-step Choquet upper/lower expectations at fixed horizons.

The Nisio Semigroup satisfies the semigroup property.

Remark 5.12 (Semigroup Property). Under Assumption 5.1, the family $(T_t)_{t \geq 0}$ is the (upper) Nisio semigroup associated with $(\bar{T}_t)_{t \geq 0}$ and satisfies the semigroup property (sometimes referred to as the dynamic programming principle),

$$T_{t+s} = T_t \circ T_s, \quad \forall t, s \geq 0.$$

To see this, we refer the reader to standard treatments of Nisio semigroups/stochastic control such as [48], Chapter III.

¹⁴The same holds for the static lower envelopes $\underline{T}_t$.

The switching construction yields a *time-consistent* nonlinear evolution: It satisfies the dynamic programming principle and acts as a coherent (sublinear) upper expectation on test functions. We summarize the resulting semigroup and regularity properties in the next proposition.

Proposition 5.13 (Sublinear Markov Semigroup on $\text{Lip}_b(E)$). *The family $\mathbf{T} = (T_t)_{t \geq 0}$ of Definition 5.11 satisfies*

- (i) $T_0 = \text{Id}$ on $\text{Lip}_b(E)$;
- (ii) $T_{t+s} = T_t \circ T_s$ for all $t, s \geq 0$;
- (iii) Each T_t is monotone, constant-preserving, sublinear, and translation invariant: $T_t(f + c) = T_t f + c$ for all $c \in \mathbb{R}$;
- (iv) $\|T_t f - T_t g\|_\infty \leq \|f - g\|_\infty$ for all $f, g \in \text{Lip}_b(E)$.

Proof. (i) For $t = 0$, Π_0 contains only $\pi = (0)$, hence $T_0^\pi f = f$ and $T_0 f = f$.

(ii) This is the semigroup property, see Remark 5.12.

(iii) Each $\bar{T}_u$ is monotone, constant-preserving, sublinear and translation invariant (as a pointwise supremum of linear Markov operators). These properties are preserved under composition (thus for T_t^π) and under pointwise suprema over π , hence they hold for T_t .

(iv) If $\|f - g\|_\infty \leq \varepsilon$, then $g - \varepsilon \leq f \leq g + \varepsilon$. By monotonicity and translation invariance, $T_t g - \varepsilon \leq T_t f \leq T_t g + \varepsilon$, yielding the claim. $\square$

To obtain long-term stability and an ergodic limit for the Nisio semigroup, we need a quantitative regularization property. The next lemma shows that the one-step Lipschitz/Wasserstein contraction propagates through switching, yielding the same contraction (hence oscillation decay) for $(T_t)_{t \geq 0}$. As a byproduct, the next lemma shows also that T_t maps $\text{Lip}_b(E)$ into $\text{Lip}_b(E)$, hence $T_t f$ is in fact Borel measurable.

Lemma 5.14 (Lipschitz contraction for $\bar{T}_t$ and T_t). *Let Assumption 5.3 hold. Then, for all $t \geq 0$ and $f \in \text{Lip}_b(E)$,*

$$\text{Lip}(\bar{T}_t f) \leq e^{-\lambda t} \text{Lip}(f), \quad \text{Lip}(T_t f) \leq e^{-\lambda t} \text{Lip}(f). \quad (5.19)$$

Consequently, under (5.1),

$$\text{osc}(T_t f) \leq D e^{-\lambda t} \text{Lip}(f). \quad (5.20)$$

Proof. Fix $t \geq 0$ and $f \in \text{Lip}_b(E)$. For each θ , Lemma 5.5 gives $\text{Lip}(P_t^\theta f) \leq e^{-\lambda t} \text{Lip}(f)$. Since the pointwise supremum of L -Lipschitz functions is L -Lipschitz,¹⁵ $\bar{T}_t f = \sup_\theta P_t^\theta f$ satisfies $\text{Lip}(\bar{T}_t f) \leq e^{-\lambda t} \text{Lip}(f)$.

For T_t , fix $\pi = (t_0, \dots, t_n) \in \Pi_t$. Iterating the previous bound along the composition defining T_t^π yields $\text{Lip}(T_t^\pi f) \leq e^{-\lambda t} \text{Lip}(f)$. Taking the supremum over π preserves the Lipschitz constant, so $\text{Lip}(T_t f) \leq e^{-\lambda t} \text{Lip}(f)$. Finally, (5.20) follows from Lemma 5.7. $\square$

We are now in a position to state the first main conclusion of this section: The time-consistent sublinear (Nisio) semigroup inherits the uniform Wasserstein/Lipschitz contraction and therefore admits a *robust* ergodic limit. In particular, for every bounded Lipschitz test functional f , the worst-case value function $T_t f$ converges uniformly to a constant (a sublinear “equilibrium expectation”), independently of the initial state.

Theorem 5.15 (Robust Ergodicity of the Nisio Semigroup Under Wasserstein Contraction). *Let Assumptions 5.1 and 5.3 hold, and assume (5.1). Then for every $f \in \text{Lip}_b(E)$ there exists a constant $\mathcal{R}(f) \in \mathbb{R}$ such that*

$$\lim_{t \rightarrow \infty} \|T_t f - \mathcal{R}(f)\|_\infty = 0.$$

¹⁵Indeed, for any $x, y \in E$, $\sup_i u_i(x) - \sup_i u_i(y) \leq \sup_i (u_i(x) - u_i(y)) \leq L d(x, y)$, and swapping x, y gives $|\sup_i u_i(x) - \sup_i u_i(y)| \leq L d(x, y)$.

Moreover, fixing $t_0 > 0$ and setting $\kappa := e^{-\lambda t_0} \in (0, 1)$, we have that, for all $t \geq 0$,

$$\|T_t f - \mathcal{R}(f)\|_\infty \leq C_{t_0} D e^{-\lambda t} \text{Lip}(f), \quad C_{t_0} := e^{\lambda t_0} \frac{2 - \kappa}{1 - \kappa}.$$

Finally, $\mathcal{R}$ is monotone, constant-preserving, sublinear, and satisfies $\mathcal{R}(T_t f) = \mathcal{R}(f)$, for all $t \geq 0$.

Proof. Fix $f \in \text{Lip}_b(E)$ and $t_0 > 0$. Set $S := T_{t_0}$ and $\kappa = e^{-\lambda t_0}$. By Lemma 5.14, $\text{osc}(S^n f) \leq D \text{Lip}(S^n f) \leq D \kappa^n \text{Lip}(f)$.

Pick $x_0 \in E$ and define $a_n := S^n f(x_0)$. Since S is monotone and constant-preserving, $\inf g \leq Sg \leq \sup g$ pointwise for every bounded g . Applying this to $g = S^n f$ yields that both a_n and a_{n+1} lie in the interval $[\inf S^n f, \sup S^n f]$; Hence,

$$|a_{n+1} - a_n| \leq \text{osc}(S^n f) \leq D \kappa^n \text{Lip}(f).$$

Thus (a_n) is Cauchy; Define $\mathcal{R}(f) := \lim_{n \rightarrow \infty} a_n$.

For any $x \in E$, $|S^n f(x) - a_n| \leq \text{osc}(S^n f) \leq D \kappa^n \text{Lip}(f)$, and therefore

$$\|S^n f - \mathcal{R}(f)\|_\infty \leq D \kappa^n \text{Lip}(f) + \sum_{k=n}^{\infty} |a_{k+1} - a_k| \leq D \kappa^n \text{Lip}(f) + \sum_{k=n}^{\infty} D \kappa^k \text{Lip}(f) = D \frac{2 - \kappa}{1 - \kappa} \kappa^n \text{Lip}(f).$$

For general $t \geq 0$, write $t = nt_0 + r$ with $r \in [0, t_0)$. By the semigroup property, $T_t f = T_r(S^n f)$. Using nonexpansiveness in $\|\cdot\|_\infty$ and constant preservation,

$$\|T_t f - \mathcal{R}(f)\|_\infty = \|T_r(S^n f) - T_r(\mathcal{R}(f))\|_\infty \leq \|S^n f - \mathcal{R}(f)\|_\infty \leq D \frac{2 - \kappa}{1 - \kappa} \kappa^n \text{Lip}(f) \leq C_{t_0} D e^{-\lambda t} \text{Lip}(f).$$

The remaining properties of $\mathcal{R}$ follow by taking limits of the maps $f \mapsto S^n f(x_0)$ and using the corresponding properties of S from Proposition 5.13. $\square$

Of course, we have a similar result for the conjugate of the Nisio semigroup, a superlinear semigroup acting as a lower expectation.

Definition 5.16 (Lower Nisio Semigroup). Let $(T_t)_{t \geq 0}$ be the upper Nisio semigroup built from the one-step envelopes $\bar{T}_t f = \sup_{\theta \in \Theta} P_t^\theta f$. Define the *lower Nisio semigroup* $(T_t^-)_{t \geq 0}$ on $\text{Lip}_b(E)$ by

$$T_t^- f := -T_t(-f), \quad t \geq 0.$$

Equivalently, with the one-step lower envelope $\underline{T}_t f := \inf_{\theta \in \Theta} P_t^\theta f$, we have

$$T_t^- f = \inf_{\pi \in \Pi_t} \left(\underline{T}_{t_n - t_{n-1}} \circ \cdots \circ \underline{T}_{t_1 - t_0} f \right),$$

interpreted as a universally measurable envelope.

Theorem 5.17 (Robust Ergodicity of the Lower Nisio Semigroup). *Let Assumptions 5.1 and 5.3 hold, and assume (5.1). Then, for every $f \in \text{Lip}_b(E)$ there exists a constant $\mathcal{R}^-(f) \in \mathbb{R}$ such that*

$$\lim_{t \rightarrow \infty} \|T_t^- f - \mathcal{R}^-(f)\|_\infty = 0.$$

Moreover, fixing $t_0 > 0$ and setting $\kappa := e^{-\lambda t_0} \in (0, 1)$, we have that for all $t \geq 0$,

$$\|T_t^- f - \mathcal{R}^-(f)\|_\infty \leq C_{t_0} D e^{-\lambda t} \text{Lip}(f), \quad C_{t_0} := e^{\lambda t_0} \frac{2 - \kappa}{1 - \kappa}.$$

Finally, $\mathcal{R}^-$ is monotone, constant-preserving, superlinear (i.e. concave):

$$\mathcal{R}^-(f + g) \geq \mathcal{R}^-(f) + \mathcal{R}^-(g), \quad \mathcal{R}^-(\alpha f) = \alpha \mathcal{R}^-(f) (\alpha \geq 0),$$

it is translation invariant $\mathcal{R}^-(f + c) = \mathcal{R}^-(f) + c$, and it satisfies $\mathcal{R}^-(T_t^- f) = \mathcal{R}^-(f)$ for all $t \geq 0$.

Proof. Let $\mathcal{R}$ be the limit functional from Theorem 5.15. By definition,

$$T_t^- f = -T_t(-f).$$

Applying Theorem 5.15 to $-f$ yields

$$\|T_t(-f) - \mathcal{R}(-f)\|_\infty \rightarrow 0$$

with the stated rate. Hence,

$$\|T_t^- f - (-\mathcal{R}(-f))\|_\infty \rightarrow 0,$$

with the same bound. Define $\mathcal{R}^-(f) := -\mathcal{R}(-f)$. All stated properties of $\mathcal{R}^-$ and $(T_t^-)_{t \geq 0}$ follow by conjugation from the corresponding sublinear properties of $\mathcal{R}$ and $(T_t)_{t \geq 0}$. $\square$

Theorems 5.15 and 5.17 give us the following reflection.

Remark 5.18 (Relationship Between Static and Pasted Limits). Let Assumptions 5.1 and 5.3 hold, and assume (5.1). Consider the upper Nisio semigroup $(T_t)_{t \geq 0}$ built from the one-step envelopes $\bar{T}_t f = \sup_{\theta \in \Theta} P_t^\theta f$ as in Definition 5.11 (and define analogously the lower Nisio semigroup $(\underline{T}_t)_{t \geq 0}$ from $\underline{T}_t f = \inf_{\theta \in \Theta} P_t^\theta f$). Then, for every $t \geq 0$ and $f \in \text{Lip}_b(E)$,

$$T_t f \geq \bar{T}_t f = \sup_{\theta \in \Theta} P_t^\theta f, \quad T_t^- f \leq \underline{T}_t f = \inf_{\theta \in \Theta} P_t^\theta f.$$

Indeed, the one-block partition $\pi = (0, t)$ belongs to Π_t , so $T_t^\pi f = \bar{T}_t f$ and $T_t f = \sup_{\pi \in \Pi_t} T_t^\pi f \geq \bar{T}_t f$; The lower bound is analogous.

Consequently, if $\mathcal{R}(f)$ and $\mathcal{R}^-(f)$ denote the ergodic limits from Theorems 5.15 (upper) and 5.17 (lower), respectively, then

$$\mathcal{R}(f) \geq \sup_{\theta \in \Theta} \mu_\theta(f), \quad \mathcal{R}^-(f) \leq \inf_{\theta \in \Theta} \mu_\theta(f),$$

where μ_θ is the unique invariant law of $\mathbf{P}^\theta$ from Proposition 5.8. In general, these inequalities may be strict, reflecting the additional flexibility of state-dependent switching allowed by the pasted (time-consistent) construction.

6. CONCLUSION

In this work, we introduced the concept of an imprecise Markov semigroup $\mathbf{Q}$, and studied its ergodic behavior (i) when the state space E is a Euclidean space or a Riemannian manifold and $\mathbf{Q}$ is compact; (ii) when E is

allowed to be an arbitrary good measurable space and $\mathbf{Q}$ is compact; and (iii) when E is a generic Polish metric space with finite diameter, and $\mathbf{Q}$ is not assumed compact.

In future work, we plan to extend the results of Section 5 beyond the function class $\text{Lip}_b(E)$ and to relax the finite-diameter assumption (5.1). We also plan to extend our theory to explicitly model ambiguity in the initial distribution of a Markov process.

Furthermore, we aim to apply these results to machine learning and computer vision, thereby making the high-level discussion in the Introduction more concrete, and to develop robust reinforcement learning methods, where Markov processes are used for planning by autonomous agents.

We are also interested in extending to the imprecise setting the types of inequalities developed in [32], Chapters 6–9. In particular, we are interested in deriving imprecise optimal transport inequalities, in the spirit of [49, 50], for application to distribution shift problems in machine learning and computer vision [51].

ACKNOWLEDGMENTS

We would like to acknowledge partial funding by the Army Research Office (ARO MURI W911NF2010080). We wish to express our gratitude to Mira Gordin for insightful discussions, particularly on how the gradient bound for a diffusion operator relates to Poincaré inequalities, and to Simone Cerreia-Vioglio and Paolo Perrone for their help in finding the suitable topology compatible with the partial order $\preceq_f^{\text{part}}$, and in shaping up the discussion after Lemma 2.9. We are also extremely grateful to Enrique Miranda for his suggestion to look into [9], Example 6.2, to Andrea Aveni for his comments on an initial version of this work, and to two anonymous reviewers for their invaluable remark that greatly improved our manuscript. We are also indebted to Malena Español for her invitation to the Institute for Advanced Study, where the idea of the present paper germinated.

REFERENCES

- [1] T. Augustin, F.P.A. Coolen, G. de Cooman and M.C.M. Troffaes, *Introduction to Imprecise Probabilities*. John Wiley & Sons, West Sussex, England (2014).
- [2] M.C.M. Troffaes and G. de Cooman, *Lower Previsions*. John Wiley and Sons, Chichester, United Kingdom (2014).
- [3] P. Walley, *Statistical Reasoning with Imprecise Probabilities*, vol. 42 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London (1991).
- [4] R.J. Crossman and D. Škulj, Imprecise Markov chains with absorption. *Int. J. Approx. Reason.* **51** (2010) 1085–1099. Imprecise probability in statistical inference and decision making.
- [5] J. De Bock, A. Erreygers and T. Krak, Sum-product laws and efficient algorithms for imprecise Markov chains, in *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, vol. 161 of *Proceedings of Machine Learning Research*, edited by C. de Campos and M.H. Maathuis. PMLR (2021) 1476–1485.
- [6] G. de Cooman, F. Hermans and E. Quaeghebeur, Imprecise Markov chains and their limit behaviour. *Probab. Eng. Inform. Sci.* **23** (2009) 597–635.
- [7] K.V. Delgado, S. Sanner and L.N. De Barros, Efficient solutions to factored MDPs with imprecise transition probabilities. *Artif. Intell.* **175** (2011) 1498–1527.
- [8] M. Jaeger, G. Bacci, G. Bacci, K.G. Larsen and P. Jensen, Approximating euclidean by imprecise Markov decision processes, in *Leveraging Applications of Formal Methods, Verification and Validation: Verification Principles: 9th International Symposium on Leveraging Applications of Formal Methods, ISoLA 2020, Rhodes, Greece, October 20–30, 2020, Proceedings, Part I*. Berlin, Heidelberg. Springer-Verlag (2020) 275–289.
- [9] T. Krak, J. De Bock and A. Siebes, Imprecise continuous-time Markov chains. *Int. J. Approx. Reason.* **88** (2017) 452–528.
- [10] T. Krak, N. T’Joens and J. De Bock, Hitting times and probabilities for imprecise Markov chains, in *Proceedings of the Eleventh International Symposium on Imprecise Probabilities: Theories and Applications*, vol. 103 of *Proceedings of Machine Learning Research*, edited by J. De Bock, C.P. de Campos, G. de Cooman, E. Quaeghebeur and G. Wheeler. PMLR (2019) 265–275.

- [11] N. T’Joens and J. De Bock, Average behaviour in discrete-time imprecise Markov chains: a study of weak ergodicity. *Int. J. Approx. Reason.* **132** (2021) 181–205.
- [12] F.W. Trevizan, F.G. Cozman and L.N. De Barros, Mixed probabilistic and nondeterministic factored planning through markov decision processes with set-valued transitions, in *Workshop on A Reality Check for Planning and Scheduling Under Uncertainty at ICAPS* (2008) 62.
- [13] F.W. Trevizan, F.G. Cozman and L.N. de Barros, Planning under risk and knightian uncertainty. In *IJCAI*, vol. 2007 (2007) 2023–2028.
- [14] M. Troffaes, T. Krak and H. Bains, Two-state imprecise Markov chains for statistical modelling of two-state non-Markovian processes, in *Proceedings of the Eleventh International Symposium on Imprecise Probabilities: Theories and Applications*, vol. 103 of *Proceedings of Machine Learning Research*, edited by J. De Bock, C.P. de Campos, G. de Cooman, E. Quaeghebeur and G. Wheeler. PMLR (2019) 394–403.
- [15] L.J.M. Aslett, F.P.A. Coolen and J. De Bock, editors, *Uncertainty in Engineering: Introduction to Methods and Applications*, vol. 7 of *SpringerBriefs in Statistics*. Springer Cham (2022).
- [16] M. Vasile, editor, *Optimization Under Uncertainty with Applications to Aerospace Engineering*, vol. 6 of *Physics and Astronomy*. Springer Cham (2021).
- [17] S. Cerreia-Vioglio, R. Corrao and G. Lanzani, Dynamic opinion aggregation: long-run stability and disagreement. *Rev. Econ. Stud.* **91** (2024) 1406–1447.
- [18] R. Denk, M. Kupper and M. Nendel, Kolmogorov-type and general extension results for nonlinear expectations. *Banach J. Math. Anal.* **12** (2018) 515–540.
- [19] R. Denk, M. Kupper and M. Nendel, A semigroup approach to nonlinear Lévy processes. *Stochastic Processes Appl.* **130** (2020) 1616–1642.
- [20] G. Choquet, Theory of capacities. *Ann. Inst. Fourier* **5** (1954) 131–295.
- [21] F. Cuzzolin, *The Geometry of Uncertainty: The Geometry of Imprecise Probabilities*. Artificial Intelligence: Foundations, Theory, and Algorithms. Springer Cham (2021).
- [22] M. Caprio and R. Gong, Dynamic precise and imprecise probability kinematics, in *Proceedings of the Thirteenth International Symposium on Imprecise Probability: Theories and Applications*, vol. 215 of *Proceedings of Machine Learning Research*, edited by E. Miranda, I. Montes, E. Quaeghebeur and B. Vantaggi. PMLR (2023) 72–83.
- [23] M. Caprio and S. Mukherjee, Ergodic theorems for dynamic imprecise probability kinematics. *Int. J. Approx. Reason.* **152** (2023) 325–343.
- [24] S. Cerreia-Vioglio, F. Maccheroni and M. Marinacci, Ergodic theorems for lower probabilities. *Proc. Am. Math. Soc.* **144** (2015) 3381–3396.
- [25] G. de Cooman, J. De Bock and S. Lopatzidis, Imprecise stochastic processes in discrete time: global models, imprecise markov chains, and ergodic theorems. *Int. J. Approx. Reason.* **76** (2016) 18–46.
- [26] F. Hermans and G. de Cooman, Characterisation of ergodic upper transition operators. *Int. J. Approx. Reason.* **53** (2012) 573–583.
- [27] I.O. Kozine and L.V. Utkin, Interval-valued finite Markov chains. *Reliable Comput.* **8** (2002) 97–113.
- [28] D. Skulj, A classification of invariant distributions and convergence of imprecise Markov chains. *Linear Algebra Appl.* **439** (2013) 3059–3076.
- [29] D. Skulj, Efficient computation of the bounds of continuous time imprecise Markov chains. *Appl. Math. Computat.* **250** (2015) 165–180.
- [30] Q. Yu, C. Li, Y. Zhu and T. Kurita, Convolutional autoencoder based on latent subspace projection for anomaly detection. *Methods* **214** (2023) 48–59.
- [31] P. Pérez, Markov random fields and images. *CWI Q.* **11** (1998) 413–437.
- [32] D. Bakry, I. Gentil and M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators*, vol. 20 of *Grundlehren der mathematischen Wissenschaften*. Springer Cham (2014).
- [33] D. Blount and M.A. Kouritzin, On convergence determining and separating classes of functions. *Stochastic Processes Appl.* **120** (2010) 1898–1907.
- [34] S.N. Ethier and T.G. Kurtz, *Markov Processes: Characterization and Convergence*. John Wiley & Sons (1986).
- [35] E.S. Wolk, Order-compatible topologies on a partially ordered set. *Proc. Am. Math. Soc.* **9** (1958) 524–529.
- [36] D. Bartl, Conditional nonlinear expectations. *Stochastic Processes Appl.* **130** (2020) 785–805.

- [37] D. Criens and L. Niemann, Markov selections and feller properties of nonlinear diffusions. *Stochastic Processes Appl.* **173** (2024) 104354.
- [38] A. Erreygers, Sublinear expectations for countable-state uncertain processes, in *Proceedings of the Thirteenth International Symposium on Imprecise Probability: Theories and Applications*, vol. 215 of *Proceedings of Machine Learning Research*, edited by E. Miranda, I. Montes, E. Quaeghebeur and B. Vantaggi. PMLR (2023) 210–221.
- [39] C. Feng and H. Zhao, Ergodicity of Sublinear Markovian Semigroups. *SIAM J. Math. Anal.* **53** (2021) 5646–5681.
- [40] F. Kühn, On infinitesimal generators of sublinear Markov semigroups. *Osaka J. Math.* **58** (2021) 487–508.
- [41] M. Nendel, Markov chains under nonlinear expectation. *Math. Finance* **31** (2021) 474–507.
- [42] M. Berger, *A Panoramic View of Riemannian Geometry*. Springer Berlin, Heidelberg (2003).
- [43] R. van Handel, Probability in high dimension. *APC 550 Lecture Notes*. , Princeton University. Available at [online](#) (2016).
- [44] R. Bishop, A relation between volume, mean curvature and diameter. *Notices Am. Math. Soc.* **10** (1963) 364.
- [45] K. Kuwada, Duality on gradient estimates and wasserstein controls. *J. Funct. Anal.* **258** (2010) 3758–3774.
- [46] G. Savaré, Self-improvement of the Bakry–Émery condition and wasserstein contraction of the heat flow in $RCD(k, \infty)$ metric measure spaces. *Discrete Continuous Dyn. Syst. A* **34** (2014) 1641–1661.
- [47] D. Azagra, J. Ferrera, F. López-Mesas and Y. Rangel, Smooth approximation of Lipschitz functions on Riemannian manifolds. *J. Math. Anal. Appl.* **326** (2007) 1370–1378.
- [48] W.H. Fleming and H. Mete Soner, *Controlled Markov Processes and Viscosity Solutions*, 2nd edn. Springer (2006).
- [49] M. Caprio, Optimal transport for ϵ -contaminated credal sets. *Theories and Applications*, PMLR **290** (2025) 33–46.
- [50] S. Lorenzini, D. Petturiti and B. Vantaggi, Choquet-wasserstein pseudo-distances via optimal transport under partially specified marginal probabilities. *Fuzzy Sets Syst.* **515** (2025) 109429.
- [51] V. Lin, K.J. Jang, S. Dutta, M. Caprio, O. Sokolsky and I. Lee, DC4L: Distribution Shift Recovery via Data-Driven Control for Deep Learning Models. Available at [arXiv:2302.10341](#) (2023).
- [52] S. Horowitz, Semi-groups of Markov Operators. *Ann. Inst. Henri Poincaré B Calcul Probab. Statist* **10** (1974) 155–166.
- [53] Y. Ito, Invariant measures for Markov processes. *Trans. Am. Math. Soc.* **110** (1964) 152–184.
- [54] E. Hille and R.S. Phillips, *Functional Analysis and Semi-Groups*. American Mathematical Society (1996).
- [55] M. Ledoux, The geometry of Markov diffusion generators. *Ann. Fac. Sci. Toulouse: Math. Ser.* **9** (2000) 305–366.
- [56] S. Lang, *Algebra*. Springer, New York (2002).

Please help to maintain this journal in open access!



This journal is currently published in open access under the Subscribe to Open model (S2O). We are thankful to our subscribers and supporters for making it possible to publish this journal in open access in the current year, free of charge for authors and readers.

Check with your library that it subscribes to the journal, or consider making a personal donation to the S2O programme by contacting subscribers@edpsciences.org.

More information, including a list of supporters and financial transparency reports, is available at <https://edpsciences.org/en/subscribe-to-open-s2o>.

APPENDIX A. PROOF OF THEOREM 4.2

We begin by pointing out that, as a consequence of Lemma 2.9 and Definition 4.1, $\underline{\mathbf{P}}$ and $\overline{\mathbf{P}}$ are (precise) Markov semigroups, and so the operators $\underline{\Gamma}$, $\overline{\Gamma}$, $\underline{L}$ and $\overline{L}$ are well-defined.

We first focus on the diffusion Markov triple $(E, \overline{\mu}, \overline{\Gamma})$. By [32], Proposition 3.3.11 and $\overline{\mathbf{P}} = (\overline{P}_t)_{t \geq 0}$ being a (precise) Markov semigroup, the fact that $\overline{L}$ is essentially self-adjoint follows from the connexity, weak hypo-ellipticity, and completeness assumptions. Pick any $f \in \mathcal{A}_0^{\overline{\mu}}$. The fact that $\overline{L}$ satisfies ESA, together with our assumption that it satisfies the curvature condition $CD(\rho_{\overline{L}}, \infty)$, for some $\rho_{\overline{L}} \in \mathbb{R}$, implies – by [32], Corollary 3.3.19 – that for all $t \geq 0$, the following

gradient bound holds

$$\bar{\Gamma}(\bar{P}_t f) \leq e^{-2t\rho_{\bar{L}}} \bar{P}_t(\bar{\Gamma}(f)). \quad (\text{A.1})$$

Then, by [32], Theorem 4.7.2, equation (A.1) implies that the following local Poincaré inequality holds for all $t \geq 0$,

$$\bar{P}_t(f^2) - (\bar{P}_t f)^2 \leq \frac{1 - e^{-2t\rho_{\bar{L}}}}{\rho_{\bar{L}}} \bar{P}_t(\bar{\Gamma}(f)). \quad (\text{A.2})$$

Here the factor needs to be understood as $2t$ when $\rho_{\bar{L}} = 0$. Following [43], Theorem 2.35, taking the limit as $t \rightarrow \infty$ on both sides of (A.2) yields the following Poincaré inequality

$$\text{Var}_{\bar{\mu}}(f) \leq \frac{1}{\rho_{\bar{L}}} \bar{\mathcal{E}}(f). \quad (\text{A.3})$$

As in the proof of Theorem 3.1, here it is implicitly assumed that $\rho_{\bar{L}} > 0$. In addition, we know that

$$\bar{\mathcal{E}}(f) = - \int_E f \bar{L} f d\bar{\mu}. \quad (\text{A.4})$$

Combining (A.4) with (A.3), we have that if function f is such that $\bar{L}f = 0$, then $\bar{\mathcal{E}}(f) = 0$. In turn, this implies that $\text{Var}_{\bar{\mu}}(f) = 0$, and so the function f must be constant $\bar{\mu}$ -a.e. Notice now that, since $\bar{L}$ satisfies the ESA property, $\mathcal{A}_0^{\bar{\mu}}$ is dense in $\mathcal{D}(\bar{\mathcal{E}})$ in the topology induced by the Dirichlet norm $\|\cdot\|_{\bar{\mathcal{E}}}$.¹⁶ Thanks to this, (A.1)–(A.4) hold for all $f \in \mathcal{D}(\bar{\mathcal{E}})$ by an approximation argument. This implies that every $f \in \mathcal{D}(\bar{\mathcal{E}})$ such that $\bar{\Gamma}(f) = 0$ is constant, and so that $\bar{L}$ is ergodic according to [32], Definition 3.1.11. Then, since $\bar{\mu}$ is finite, by [32], Proposition 3.1.13 we have that

$$\forall f \in \mathbb{L}^2(\bar{\mu}), \quad \lim_{t \rightarrow \infty} \bar{P}_t f = \int_E f d\bar{\mu}, \quad \bar{\mu}\text{-a.e.}$$

in $\mathbb{L}^2(\bar{\mu})$. This concludes the first part of the proof. The second part, that is, the one pertaining the least element $\underline{\mathbf{P}}$ of $\mathbf{Q}$, is analogous. The last statement of the theorem is immediate once we let $\bar{\mu} = \underline{\mu} = \mu$. $\square$

APPENDIX B. MEASURABILITY OF ENVELOPES AND SWITCHING OPERATORS

Under Assumption 5.1, for each $t \geq 0$ and bounded Borel f , the map $(x, \theta) \mapsto P_t^\theta f(x)$ is Borel measurable (by a standard monotone class argument, starting from indicators $\mathbb{1}_A$). Now, let us denote by $\mathcal{F}^*$ the *universal completion* of $\mathcal{F}$, and by $B^*(E)$ the space of bounded $\mathcal{F}^*$ -measurable real-valued functionals on E . Since f is $\mathcal{F}^*$ -measurable, it is measurable with respect to the completion of $p_t^\theta(x, \cdot)$ for every (t, x, θ) ; Hence we may extend P_t^θ to $B^*(E)$ by

$$P_t^\theta f(x) := \int_E f(y) p_t^\theta(x, dy), \quad f \in B^*(E),$$

which is well-defined and $\mathcal{F}^*$ -measurable in x .

Consequently, for each fixed $t \geq 0$ and bounded Borel f , the envelope $x \mapsto \sup_{\theta \in \Theta} P_t^\theta f(x)$ is upper semianalytic, hence $\mathcal{F}^*$ -measurable. Likewise, the *time-consistent* switching operator introduced in Definition 5.11 is obtained by composing the one-step envelopes $\bar{T}$ along a finite partition of $[0, t]$ and then taking a supremum over partitions.¹⁷ Since upper semianalytic functions are stable under application of Borel kernels – i.e. $g \mapsto (x \mapsto \int g(y) p(x, dy))$ – and under pointwise suprema, for bounded Borel f each $T_t f$ is upper semianalytic and therefore $\mathcal{F}^*$ -measurable. For general $f \in B^*(E)$ the same conclusion holds at the level of universal measurability, so all operators are interpreted as acting on $B^*(E)$.

¹⁶This norm is recalled in Definition C.9.(DMT4) in C.2.

¹⁷Similarly for the lower Nisio semigroup in Definition 5.16.

APPENDIX C. BACKGROUND NOTIONS

In an effort to keep the paper self-contained, we present in this section the preliminary concepts needed to appreciate the main results in our work.

C.1 Background on Markov semigroups

Let us briefly recall what a (precise) Markov process is; To do so, we borrow the notation and terminology from [32], Chapter 1.1. Let $(X_t)_{t \geq 0}$ be a measurable process on a probability space $(\Omega, \Sigma, \mathbb{P})$, such that $X_t \equiv X_t(\omega)$ is an element of the good measurable state space E from Definition 2.1, for all $t \geq 0$ and all $\omega \in \Omega$. Denote by $\mathcal{F}_t := \sigma(X_u : u \leq t)$, $t \geq 0$, the natural filtration of $(X_t)_{t \geq 0}$.

The *Markov property* indicates that for $t > s$, the law of X_t given $\mathcal{F}_s$ is the law of X_t given X_s , as well as the law of X_{t-s} given X_0 , the latter property reflecting the fact that the Markov process is time homogeneous, which is the only case that we consider in the present work. As a consequence, throughout the paper we focus our attention on the law of X_t given X_0 , for a generic $t \geq 0$.

For any $t \geq 0$, we can describe the distribution of X_t starting from $X_0 = x$ via a *probability kernel* $p_t(x, A)$, that is, a function $p_t : E \times \mathcal{F} \rightarrow [0, 1]$ such that (i) $p_t(x, \cdot)$ is a probability measure, for all $x \in E$, and (ii) $x \mapsto p_t(x, A)$ is measurable, for every measurable set $A \in \mathcal{F}$.

We now introduce one of the main concepts that we use in the present paper, namely a *Markov semigroup*. We follow the presentation in [32], Chapter 1.2. Let $B(E)$ denote the set of bounded measurable functionals on $(E, \mathcal{F})$. A Markov semigroup $\mathbf{P} = (P_t)_{t \geq 0}$ is a family of operators that satisfies the following properties.

- (i) For every $t \geq 0$, $P_t : B(E) \rightarrow B(E)$ is a linear operator (linearity).
- (ii) $P_0 = \text{Id}$, the identity operator (initial condition).
- (iii) $P_t \mathbb{1} = \mathbb{1}$, where $\mathbb{1}$ is the constant function equal to 1 (mass conservation).¹⁸
- (iv) For every $t \geq 0$, $f \geq 0 \implies P_t f \geq 0$ (positivity preserving).
- (v) For every $t, s \geq 0$, $P_{t+s} = P_t \circ P_s$ (semigroup property).

One extra condition – continuity at $t = 0$ – is needed to fully define a Markov semigroup $\mathbf{P} = (P_t)_{t \geq 0}$. To present it though, we first need to introduce the concept of invariant (or stationary) measure.

Let $\mathbf{P} = (P_t)_{t \geq 0}$ be a family of operators satisfying (i)–(v). Then, a (positive) σ -finite measure μ on $(E, \mathcal{F})$ is called *invariant* for $\mathbf{P}$ if for every bounded positive measurable functional f on E , and every $t \geq 0$,

$$\int_E P_t f d\mu = \int_E f d\mu. \quad (\text{C.1})$$

Notice that, as a consequence of (iii), (iv), and Jensen's inequality, we have that if a measure μ is invariant for $\mathbf{P}$, then, for every $t \geq 0$, P_t is a contraction on the bounded functions in $\mathbb{L}^p(\mu)$, for any $1 \leq p \leq \infty$.

We are now ready for the complete definition.

Definition C.1 (Markov Semigroup). Let $(E, \mathcal{F})$ be a good measurable space. A family $\mathbf{P} = (P_t)_{t \geq 0}$ of operators P_t is called a Markov semigroup if it satisfies (i)–(v), there exists an invariant measure μ for $\mathbf{P}$, and the following extra condition is satisfied

- (vi) For every $f \in \mathbb{L}^2(\mu)$, $\lim_{t \rightarrow 0} P_t f = f$ in $\mathbb{L}^2(\mu)$ (continuity at $t = 0$).

Notice that, since $(P_t)_{t \geq 0}$ is a contraction (as pointed out after equation (C.1)) and a semigroup (by condition (v)), property (vi) expresses that $t \mapsto P_t f$ is continuous in $\mathbb{L}^2(\mu)$ on $\mathbb{R}_+$. In addition, we could have required convergence in $\mathbb{L}^p(\mu)$ for any $p \in [1, \infty)$, but we stick to $p = 2$ for simplicity and to align with [32]. Before moving on, let us add a remark on invariant measures.

Remark C.2 (Spotlight on Invariant Measures). In general, an invariant measure μ for a Markov semigroup $\mathbf{P}$ need not be a *probability* measure.¹⁹ In addition, it need not always exist; Necessary and sufficient conditions for its existence may be found in [52, 53]. That being said, most semigroups of interest do have an invariant measure [32], Section 1.2.1.

¹⁸In general, we write $P_t(f) \equiv P_t f$, for all $f \in B(E)$, for notational convenience.

¹⁹If it is a finite measure, it is customary to normalize it to a probability measure. It is then usually unique [32], p. 11.

A way to derive μ as a weak limit of a “reasonable” initial probability measure μ_0 is inspected in [32], p. 10. A typical example of a Markov semigroup having an invariant measure is the heat or Brownian semigroup on $\mathbb{R}^n$, whose invariant measure is (up to multiplication by a constant) the Lebesgue measure.²⁰

As pointed out in [32], Section 1.2.1, when the invariant measure μ is a probability measure, it has an immediate interpretation. If the Markov process $(X_t)_{t \geq 0}$ starts at time $t = 0$ with initial distribution μ , *i.e.* if $X_0 \sim \mu$, then it keeps the same distribution at each time $t \geq 0$; Indeed,

$$\begin{aligned}\mathbb{E}[f(X_t)] &= \mathbb{E}[\mathbb{E}(f(X_t) \mid X_0)] = \mathbb{E}[P_t f(X_0)] = \int_E P_t f d\mu \\ &= \int_E f d\mu = \mathbb{E}[f(X_0)],\end{aligned}$$

for all $f \in B(E)$ and all $t \geq 0$. Functionals f , then, have to be understood as classes of functions for the μ -a.e. equality, and equalities and inequalities such as $f \leq g$ are always understood to hold μ -a.e.

We now continue our presentation of the concepts related to Markov semigroups that are crucial to derive our main results. An important property of a Markov semigroup is its symmetry [32], Definition 1.6.1.

Definition C.3 (Symmetric Markov Semigroup). Let $\mathbf{P} = (P_t)_{t \geq 0}$ be a Markov semigroup with (good) state space $(E, \mathcal{F})$ and invariant measure μ . We say $\mathbf{P}$ is *symmetric* with respect to μ – or equivalently, that μ is *reversible* for $\mathbf{P}$ – if, for all functions $f, g \in \mathbb{L}^2(\mu)$ and all $t \geq 0$,

$$\int_E f P_t g d\mu = \int_E g P_t f d\mu.$$

The probabilistic interpretation of Definition C.3 is straightforward. As pointed out in [32], Section 1.6.1, the name “reversible” refers to reversibility in time of the Markov process $(X_t)_{t \geq 0}$ associated with the Markov semigroup $\mathbf{P}$ whenever the initial law μ is the invariant measure. Indeed, from (C.1), we know that if the process starts from the invariant distribution μ , then it keeps the same distribution at each time t . Moreover, if the measure is reversible, and if the initial distribution of X_0 is μ , then for any $t > 0$ and any partition $0 \leq t_1 \leq \dots \leq t_k \leq t$ of the time interval $[0, t]$, the law of $(X_0, X_{t_1}, \dots, X_{t_k}, X_t)$ is the same as the law of $(X_t, X_{t-t_1}, \dots, X_{t-t_k}, X_0)$. Hence, we can say that the law of the Markov process is “reversible in time”.²¹

We now need to introduce the concept of infinitesimal generator of a Markov semigroup $\mathbf{P} = (P_t)_{t \geq 0}$ [32], Definition 1.4.1. By Hille-Yosida theory [54],²² there exists a dense linear subspace of $\mathbb{L}^2(\mu)$, called the domain $\mathcal{D}$ of the semigroup $\mathbf{P}$, on which the derivative at $t = 0$ of P_t exists in $\mathbb{L}^2(\mu)$.

Definition C.4 (Infinitesimal Generator). Let $\mathbf{P} = (P_t)_{t \geq 0}$ be a Markov semigroup with (good) state space $(E, \mathcal{F})$ and invariant measure μ . The *infinitesimal generator* L of $\mathbf{P}$ in $\mathbb{L}^2(\mu)$ is a linear operator $L : \mathcal{D} \rightarrow \mathbb{L}^2(\mu)$,

$$f \mapsto Lf := \lim_{t \rightarrow 0} \frac{P_t f - f}{t}.$$

The domain $\mathcal{D}$ is also called the domain of L , denoted by $\mathcal{D}(L)$, and depends on the underlying space $\mathbb{L}^2(\mu)$. Two comments are in place. First, as indicated by [32], Chapter 1.4, the couple $(L, \mathcal{D}(L))$ completely characterizes the Markov semigroup $\mathbf{P}$ acting on $\mathbb{L}^2(\mu)$. Second, the infinitesimal generator L has a natural interpretation. As pointed out by Dan MacKinlay, L is a kind of linearization of the local Markov transition kernel for the Markov process $(X_t)_{t \geq 0}$ associated with the Markov semigroup $\mathbf{P}$; This is because L is defined as a derivative.

The infinitesimal generator L of a Markov semigroup $\mathbf{P}$ is sometimes referred to as the *Markov generator* (of $\mathbf{P}$). We can express the reversibility property of Definition C.3 in terms of L as

$$\int_E f L g d\mu = \int_E g L f d\mu, \quad \forall f, g \in \mathcal{D}(L). \quad (\text{C.2})$$

²⁰The invariant measure is in general only defined up to a multiplicative constant.

²¹For examples of symmetric Markov semigroups, we refer the reader to [32], Section 1.6.1, Appendices A.2, A.3.

²²A more modern reference for this result is [32], Appendix A.1.

The Markov generator is thus symmetric, however it is unbounded in $\mathbb{L}^2(\mu)$.

Next, we present the carré du champ operator, that measures how far an infinitesimal generator L is from being a derivation [55].²³

Definition C.5 (Carré du Champ). Let $\mathbf{P} = (P_t)_{t \geq 0}$ be a Markov semigroup with (good) state space $(E, \mathcal{F})$ and invariant measure μ . Let L be its infinitesimal generator with $\mathbb{L}^2(\mu)$ -domain $\mathcal{D}(L)$. Let $\mathcal{A} \subseteq \mathcal{D}(L)$ be an algebra. Then, the bilinear map

$$\Gamma : \mathcal{A} \times \mathcal{A} \rightarrow \mathcal{A}, \quad (f, g) \mapsto \Gamma(f, g) := \frac{1}{2} [L(fg) - fLg - gLf]$$

is called the *carré du champ operator* of the Markov generator L .

The carré du champ operator is symmetric and is positive on $\mathcal{A}$, that is, $\Gamma(f, f) \geq 0$, for all $f \in \mathcal{A}$.²⁴ In this work, we write $\Gamma(f, f) \equiv \Gamma(f)$, $f \in \mathcal{A}$, to lighten the notation. In addition, notice that the definition of the carré du champ operator is subordinate to the algebra $\mathcal{A}$. For this reason, we call the latter the *carré du champ algebra*.

We now ask ourselves whether it is enough to know the Markov generator L on smooth functionals on E in order to describe the associated Markov semigroup $\mathbf{P} = (P_t)_{t \geq 0}$. The answer is positive, provided L is a diffusion operator [32], Definition 1.11.1.

Definition C.6 (Diffusion Operator). An operator L , with carré du champ operator Γ , is said to be a *diffusion operator* if

$$L\psi(f) = \psi'(f)Lf + \psi''(f)\Gamma(f), \quad (\text{C.3})$$

for all $\psi : \mathbb{R} \rightarrow \mathbb{R}$ of class at least $\mathcal{C}^2$, and every suitably smooth functional f on E .²⁵

Together with the following [32], Section 1.16.1, the notion of diffusion operator will allow us to link the probabilistic properties of a Markov semigroup to the geometry of the state space E .

Definition C.7 (Iterated Carré du Champ Operator). Let $\mathbf{P} = (P_t)_{t \geq 0}$ be a symmetric Markov semigroup with (good) state space $(E, \mathcal{F})$ and invariant reversible measure μ . Let L be its infinitesimal generator with $\mathbb{L}^2(\mu)$ -domain $\mathcal{D}(L)$, and denote by $\mathcal{A}$ the algebra on which the carré du champ operator Γ of L is well defined. The *iterated carré du champ operator* of the Markov generator L is defined as

$$\mathcal{A} \ni \Gamma_2(f, g) := \frac{1}{2} [L\Gamma(f, g) - \Gamma(f, Lg) - \Gamma(Lf, g)], \quad (\text{C.4})$$

for every pair $(f, g) \in \mathcal{A} \times \mathcal{A}$ such that the terms on the right hand side of (C.4) are well defined.

As for the carré du champ operator Γ , we write $\Gamma_2(f, f) \equiv \Gamma_2(f)$ to lighten the notation. Unlike Γ , Γ_2 is not always positive. As we will see more formally in C.1.1, Γ_2 is positive if and only if the state space E has “positive curvature”. This high-level intuition, together with the fact that many differential inequalities can be expressed *via* Γ_2 and Γ , justifies the following definition [32], Definition 1.16.1.

Definition C.8 (Bakry-Émery Curvature). Let $\mathbf{P} = (P_t)_{t \geq 0}$ be a symmetric Markov semigroup with (good) state space $(E, \mathcal{F})$ and invariant reversible measure μ . Let L be its infinitesimal generator with $\mathbb{L}^2(\mu)$ -domain $\mathcal{D}(L)$, and assume it is a diffusion operator. Denote by $\mathcal{A}$ the algebra on which the carré du champ operator Γ of L is well defined.²⁶ Operator L is said to satisfy the *Bakry-Émery curvature condition* $CD(\rho, n)$, for $\rho \in \mathbb{R}$ and $n \in [1, \infty]$, if for every functional f in $\mathcal{A}$, the following holds μ -a.e.

$$\Gamma_2(f) \geq \rho\Gamma(f) + \frac{1}{n}(Lf)^2. \quad (\text{C.5})$$

²³Here, “derivation” refers to the differential algebra meaning of the term [56].

²⁴This is a fundamental property of Markov semigroups [32], Section 1.4.2.

²⁵Typically of class at least $\mathcal{C}^2$.

²⁶e.g. the algebra of functionals of class at least $\mathcal{C}^2$ on E .

Of course, equation (C.5) becomes $\Gamma_2(f) \geq \rho\Gamma(f)$ for $n = \infty$.

C.1.1 Riemannian manifold

When $E = \mathbb{R}^n$, or $E = M$, a complete (with respect to the Riemannian distance) Riemannian manifold (see Fig. 3), the carré du champ and the iterated carré du champ operators are easily computed.

For example, if $\mathcal{A}_0$ is the class of smooth ($\mathcal{C}^\infty$) functionals with compact support²⁷ – i.e. such that $\text{supp}(f) = \{x \in E : f(x) \neq 0\}$ is compact in the Euclidean topology – and $L = \Delta$ (the standard Laplacian on $\mathbb{R}^n$), then $\Gamma(f) = |\nabla f|^2$ and $\Gamma_2(f) = |\nabla \nabla f|^2$, where $\nabla \nabla f = \text{Hess}(f) = (\partial_{ij} f)_{i,j \in \{1, \dots, n\}}$ is the Hessian of f . Let instead L be the Laplace-Beltrami operator $\Delta_{\mathbf{g}}$ on an n -dimensional Riemannian manifold $(M, \mathbf{g})$. Here $\mathbf{g}$ is the Riemannian co-metric $\mathbf{g} \equiv \mathbf{g}(x) = (g^{ij})_{i,j \in \{1, \dots, n\}} = ((g_{ij})_{i,j \in \{1, \dots, n\}})^{-1} = G^{-1}$, where G is the positive-definite symmetric matrix that gives the Riemannian metric. Then, $\Gamma(f) = |\nabla f|^2$ and – thanks to the Bochner-Lichnerowicz formula [32], Theorem C.3.3 – $\Gamma_2(f) = |\nabla \nabla f|^2 + \text{Ric}(\nabla f, \nabla f)$, where $\text{Ric} = \text{Ric}_{\mathbf{g}}$ is the Ricci tensor of $(M, \mathbf{g})$.²⁸ This implies that

$$\Gamma_2(f) \geq \rho\Gamma(f) = \rho|\nabla f|^2, \forall f \in \mathcal{A}_0 \iff \text{Ric}(\nabla f, \nabla f) \geq \rho|\nabla f|^2,$$

that is, the Bakry-Émery curvature condition $CD(\rho, \infty)$ holds if and only if the Ricci tensor at every point is bounded from below by ρ . In turn, Γ_2 is positive if and only if the Ricci curvature Ric of the manifold is positive. This makes more precise the idea we expressed in the previous section that Γ_2 is positive if and only if the state space E has “positive curvature”.

We now introduce the concept of elliptic diffusion operator. First, notice that when E is $\mathbb{R}^n$ or a Riemannian manifold M , we can rewrite condition (C.3) for L to be a diffusion operator as

$$Lf = \sum_{i,j \in \{1, \dots, n\}} g^{ij} \partial_{ij}^2 f + \sum_{i=1}^n b^i \partial_i f, \quad f \in \mathcal{A}_0. \quad (\text{C.6})$$

Here, $\mathbf{g} \equiv \mathbf{g}(x) = (g^{ij}(x))_{i,j \in \{1, \dots, n\}}$ is a smooth ($\mathcal{C}^\infty$) $n \times n$ symmetric matrix-valued function of $x \in E$ (e.g. the co-metric of M at x), and $b \equiv b(x) = (b^i(x))_{i=1}^n$ is a smooth ($\mathcal{C}^\infty$) $\mathbb{R}^n$ -valued function of $x \in E$ (e.g. a vector field). In turn, this implies that we can write the carré du champ operator as $\Gamma(f, g) = \sum_{i,j \in \{1, \dots, n\}} g^{ij} \partial_i f \partial_j g$, $f, g \in \mathcal{A}_0$. A diffusion operator as in (C.6) is said to be *elliptic* if

$$\mathbf{g}(V, V) = \sum_{i,j \in \{1, \dots, n\}} g^{ij}(x) V_i V_j \geq 0, \quad \forall V = (V_i)_{i=1}^n \in \mathbb{R}^n$$

and

$$\mathbf{g}(V, V) = 0 \implies V = 0.$$

Loosely, the fact that a diffusion operator L is elliptic is important because it ensures that any solution on $[0, s] \times \mathcal{O}$ of the heat equation $\partial_t u = Lu$, $u(0, x) = u_0$, where $s \in \mathbb{R}_{>0}$ and $\mathcal{O} \subseteq \mathbb{R}^n$ is open in the Euclidean topology, is $\mathcal{C}^\infty$ on $(0, s) \times \mathcal{O}$, no matter the initial condition u_0 .

C.2 Background on Diffusion Markov Triples

We first introduce Diffusion Markov Triples.

Definition C.9 (Diffusion Markov Triple). A Diffusion Markov Triple (E, μ, Γ) is composed of a measure space $(E, \mathcal{F}, \mu)$, a class $\mathcal{A}_0$ of bounded measurable functionals on E , and a symmetric bilinear operator (the carré du champ operator) $\Gamma : \mathcal{A}_0 \times \mathcal{A}_0 \rightarrow \mathcal{A}_0$, satisfying properties D1-D9 in [32], Section 3.4.1. We report them here to keep the paper self-contained.

²⁷The reason for subscript 0 in $\mathcal{A}_0$ will be clear in Section C.2.

²⁸Loosely, the Ricci tensor gives us a measure of how the geometry of a given metric tensor differs locally from that of ordinary Euclidean space. Intuitively, it quantifies how a shape is deformed as one moves along the geodesics in the Riemannian manifold.

- (DMT1) Measure space $(E, \mathcal{F}, \mu)$ is a *good measure space*, that is, there is a countable family of sets which generates $\mathcal{F}$ (up to sets of μ -measure 0), and both the decomposition and bi-measure theorems apply. Measure μ is σ -finite and, when finite, assumed to be a probability.
- (DMT2) $\mathcal{A}_0$ is a vector space of bounded measurable functionals on E , which is dense in every $\mathbb{L}^p(\mu)$, $1 \leq p < \infty$, stable under products (that is, $\mathcal{A}_0$ is an algebra) and stable under the action of smooth ($\mathcal{C}^\infty$) functions $\Psi : \mathbb{R}^k \rightarrow \mathbb{R}$ vanishing at 0, *i.e.* such that $\Psi(0) = 0$.
- (DMT3) The carré du champ operator Γ is positive, that is, $\Gamma(f) \equiv \Gamma(f, f) \geq 0$, for all $f \in \mathcal{A}_0$.²⁹ It satisfies the following diffusion hypothesis: for every smooth ($\mathcal{C}^\infty$) function Ψ vanishing at 0, and every $f_1, \dots, f_k, g \in \mathcal{A}_0$,

$$\Gamma(\Psi(f_1, \dots, f_k), g) = \sum_{i=1}^k \partial_i \Psi(f_1, \dots, f_k) \Gamma(f_i, g). \quad (\text{C.7})$$

- (DMT4) For every $f \in \mathcal{A}_0$, there exists a finite constant $C(f)$ such that for every $g \in \mathcal{A}_0$,

$$\left| \int_E \Gamma(f, g) d\mu \right| \leq C(f) \|g\|_2.$$

The Dirichlet form $\mathcal{E}$ is defined for every $(f, g) \in \mathcal{A}_0 \times \mathcal{A}_0$ by

$$\mathcal{E}(f, g) = \int_E \Gamma(f, g) d\mu,$$

and we write $\mathcal{E}(f, f) \equiv \mathcal{E}(f)$ for notational convenience. The domain $\mathcal{D}(\mathcal{E})$ of the Dirichlet form $\mathcal{E}$ is the completion of $\mathcal{A}_0$ with respect to the norm $\|f\|_{\mathcal{E}} = [\|f\|_2^2 + \mathcal{E}(f)]^{1/2}$. The Dirichlet form $\mathcal{E}$ is extended to $\mathcal{D}(\mathcal{E})$ by continuity together with the carré du champ operator Γ .

- (DMT5) L is a linear operator on $\mathcal{A}_0$ defined by and satisfying the integration by parts formula

$$\int_E g Lf d\mu = - \int_E \Gamma(f, g) d\mu,$$

for all $f, g \in \mathcal{A}_0$. A consequence of condition (DMT3) is that

$$\begin{aligned} L(\Psi(f_1, \dots, f_k)) &= \sum_{i=1}^k \partial_i \Psi(f_1, \dots, f_k) Lf_i \\ &\quad + \sum_{i,j=1}^k \partial_{ij}^2 \Psi(f_1, \dots, f_k) \Gamma(f_i, f_j), \end{aligned} \quad (\text{C.8})$$

for a smooth ($\mathcal{C}^\infty$) functions $\Psi : \mathbb{R}^k \rightarrow \mathbb{R}$ such that $\Psi(0) = 0$, and $f_1, \dots, f_k \in \mathcal{A}_0$. Hence L too satisfies the diffusion property as we introduced it in Definition C.6.

- (DMT6) For the operator L defined in (DMT5), $L(\mathcal{A}_0) \subset \mathcal{A}_0$, *i.e.* $Lf \in \mathcal{A}_0$, for all $f \in \mathcal{A}_0$.
- (DMT7) The domain $\mathcal{D}(L)$ of the operator L is defined as the set of $f \in \mathcal{D}(\mathcal{E})$ for which there exists a finite constant $C(f)$ such that, for any $g \in \mathcal{D}(\mathcal{E})$,

$$|\mathcal{E}(f, g)| \leq C(f) \|g\|_2.$$

On $\mathcal{D}(L)$, L is extended *via* the integration by parts formula for every $g \in \mathcal{D}(\mathcal{E})$.

- (DMT8) For every $f \in \mathcal{A}_0$, $\int_E Lf d\mu = 0$.
- (DMT9) The semigroup $\mathbf{P} = (P_t)_{t \geq 0}$ is the symmetric semigroup with infinitesimal generator L defined on its domain $\mathcal{D}(L)$. In general, it is sub-Markov, that is, $P_t(\mathbb{1}) \leq \mathbb{1}$, for all $t \geq 0$, where $\mathbb{1}$ is the constant function equal to 1. For it to be Markov, $P_t(\mathbb{1}) = \mathbb{1}$ must hold for all $t \geq 0$. Moreover, $\mathcal{A}_0$ is not necessarily dense in the domain

²⁹The carré du champ operator Γ also satisfies [32], Equation (3.1.1), that is, $\int_E \Gamma(g, f^2) d\mu + 2 \int_E g \Gamma(f) d\mu = 2 \int_E \Gamma(fg, f) d\mu$.

$\mathcal{D}(L)$ with respect to the domain norm $\|f\|_{\mathcal{D}(L)} = [\|f\|_2^2 + \|Lf\|_2^2]^{1/2} \equiv \|f\|_{\mathbb{L}^2(\mu)} + \|Lf\|_{\mathbb{L}^2(\mu)}$. By (DMT8), we have that μ is invariant for $\mathbf{P}$. In addition, from the symmetry of Γ we have that μ is reversible for $\mathbf{P}$.

Given Definition C.9, we see how L is a second order differential operator, Γ is a first order operator in each of its arguments, and Γ_2 (see Def. C.7) is a second order differential in each of its arguments.

Before recalling adjoints and essential self-adjointness, we briefly fix notation on the graph norm and closedness for unbounded operators, since we will use approximation on $\mathcal{D}(L)$ through a core.

Let $(H, \|\cdot\|_H)$ be a Banach space and let $A : \mathcal{D}(A) \subset H \rightarrow H$ be a linear operator. The *graph* of A is

$$\text{Graph}(A) := \{(f, Af) : f \in \mathcal{D}(A)\} \subset H \times H,$$

and the associated *graph norm* on $\mathcal{D}(A)$ is

$$\|f\|_{\mathcal{D}(A)} := \|f\|_H + \|Af\|_H.$$

We recall that A is *closed* if and only if $\text{Graph}(A)$ is closed in $H \times H$. Equivalently, A is closed if and only if $(\mathcal{D}(A), \|\cdot\|_{\mathcal{D}(A)})$ is a Banach space. Indeed, the map $f \mapsto (f, Af)$ is an isometric embedding of $\mathcal{D}(A)$ into $H \times H$ endowed with $\|(u, v)\|_{H \times H} := \|u\|_H + \|v\|_H$, and completeness is equivalent to closedness of the range. In our setting, $H = \mathbb{L}^2(\mu)$ and the generator L of the strongly continuous Markov semigroup $(P_t)_{t \geq 0}$ is closed; Hence, $\mathcal{D}(L)$ is complete for the graph norm $\|\cdot\|_{\mathcal{D}(L)}$.

When we write that $\mathcal{A}_0$ is dense in $\mathcal{D}(L)$ for the graph norm, we mean that $\mathcal{A}_0$ is a *core* for L : For every $f \in \mathcal{D}(L)$ there exists $(f_n)_n \subset \mathcal{A}_0$ such that

$$\|f_n - f\|_{\mathbb{L}^2(\mu)} + \|Lf_n - Lf\|_{\mathbb{L}^2(\mu)} \rightarrow 0.$$

This approximation is precisely what allows us to extend identities first proved on $\mathcal{A}_0$ (e.g. integration by parts and the Poincaré inequality) to all $f \in \mathcal{D}(L)$ by continuity.

We now introduce the concepts of adjoint operator, self-adjointness and essential self-adjointness (ESA) [32], Section 3.4.2. The latter allow us in Theorem 4.2, to relax the requirement in Theorem 3.1 that $\underline{L}$ and $\bar{L}$ are elliptic operators.

Definition C.10 (Adjoint Operator). Let (E, μ, Γ) be a diffusion Markov triple, and consider the operator L associated with it. The domain $\mathcal{D}(L^*)$ is the set of functions $f \in \mathbb{L}^2(\mu)$ such that there exists a finite constant $C(f)$ for which, for all $g \in \mathcal{A}_0$,

$$\left| \int_E f L g d\mu \right| \leq C(f) \|g\|_2.$$

On this domain, the *adjoint operator* L^* is defined by integration by parts: For any $g \in \mathcal{A}_0$,

$$L^*(f) = \int_E f L g d\mu = \int_E g L^* f d\mu.$$

By symmetry of L – as we have seen in (C.2) – it holds that (in general) $\mathcal{D}(L) \subset \mathcal{D}(L^*)$, and $L = L^*$ on $\mathcal{D}(L)$. That is, L is *self-adjoint* on $\mathcal{D}(L)$. Being self-adjoint on $\mathcal{D}(L)$, though, does not mean that $L = L^*$. The self-adjointness property of L refers to the dual operator constructed in the same way as L^* , but where $\mathcal{D}(L)$ replaces $\mathcal{A}_0$. In other words, as we have seen in Definition C.9.(DMT9), $\mathcal{A}_0$ is not necessarily dense in $\mathcal{D}(L)$ with respect to the domain norm $\|\cdot\|_{\mathcal{D}(L)}$. This happens if and only if $\mathcal{D}(L) = \mathcal{D}(L^*)$. In this case, the extension of L from $\mathcal{A}_0$ to a larger domain as a self-adjoint operator is unique.

Definition C.11 (Essential Self-Adjointness – ESA). The operator L is said to be essentially self-adjoint if $\mathcal{D}(L) = \mathcal{D}(L^*)$ (recall that $\mathcal{D}(L^*)$ is defined with respect to $\mathcal{A}_0$). Equivalently, $\mathcal{A}_0$ is dense in $\mathcal{D}(L)$ in the topology induced by $\|\cdot\|_{\mathcal{D}(L)}$.

As we can see, the ESA property depends on the choice of $\mathcal{A}_0$. In addition, ESA holds for all elliptic operators on complete Riemannian manifolds where $\mathcal{A}_0$ is the class of smooth compactly supported functions. That is, it is always

satisfied when E is a Euclidean space or a complete Riemannian manifold. The other implication, though, need not hold – that is, there are ESA operators that are not elliptic as in C.1.1. As we shall see, ESA alone is enough to guarantee ergodicity à la Corollary 3.3.

The ESA property does not automatically hold for a diffusion Markov triple (E, μ, Γ) . To derive sufficient conditions for it, we need to introduce the concept of an extension of $\mathcal{A}_0$ [32], Section 3.3.1.

Definition C.12 (Extended Algebra $\mathcal{A}$). Let $\mathcal{A}_0$ be the algebra associated with a diffusion Markov triple (E, μ, Γ) . $\mathcal{A}$ is an algebra of measurable functions on E containing $\mathcal{A}_0$, containing the constant functions, and satisfying the following requirements.

- (i) For all $f \in \mathcal{A}$ and all $h \in \mathcal{A}_0$, $hf \in \mathcal{A}_0$.
- (ii) For any $f \in \mathcal{A}$, if $\int_E hf d\mu \geq 0$ for all positive $h \in \mathcal{A}_0$, then $f \geq 0$.
- (iii) $\mathcal{A}$ is stable under composition with smooth (C^∞) functions $\Psi : \mathbb{R}^k \rightarrow \mathbb{R}$.
- (iv) The operator $L : \mathcal{A} \rightarrow \mathcal{A}$ is an extension of the infinitesimal generator L on $\mathcal{A}_0$. The carré du champ operator Γ is also defined on $\mathcal{A} \times \mathcal{A}$ as in Definition C.5, with respect to the extension of L .
- (v) For every $f \in \mathcal{A}$, $\Gamma(f) \geq 0$.
- (vi) The operators Γ and L satisfy (C.7) and (C.8), respectively, for any smooth (C^∞) functions $\Psi : \mathbb{R}^k \rightarrow \mathbb{R}$.
- (vii) For every $f \in \mathcal{A}$ and every $g \in \mathcal{A}_0$, the following holds

$$\mathcal{E}(f, g) = \int_E \Gamma(f, g) d\mu = - \int_E g Lf d\mu = - \int_E f Lg d\mu.$$

- (viii) For every $f \in \mathcal{A}_0$ and every $t \geq 0$, $P_t f \in \mathcal{A}$.

When E is a Euclidean space or a complete Riemannian manifold, we have that $\mathcal{A}_0$ is the class of smooth (C^∞) compactly supported functionals on E , while $\mathcal{A}$ is the class of smooth (C^∞) functionals. We now present two concepts, connexity and weak hypo-ellipticity, that allow to “import” some of the geometric structure of the Riemannian manifolds considered in C.1.1 to the more general diffusion Markov triple setting [32], Section 3.3.3. In addition, as we can see in the proof of Theorem 4.2, in the context of imprecise diffusion Markov tuples they are sufficient for the ESA property to hold.

Definition C.13 (Connexity and Weak Hypo-Ellipticity). Let (E, μ, Γ) be a diffusion Markov triple, and let $\mathcal{A}$ be the extension of the algebra $\mathcal{A}_0$ associated with (E, μ, Γ) . Then, (E, μ, Γ) is said to be *connected* if $f \in \mathcal{A}$ and $\Gamma(f) = 0$ imply that f is constant. This is a local property for functionals in $\mathcal{A}$. (E, μ, Γ) is *weakly hypo-elliptic* if, for every $\lambda \in \mathbb{R}$, any $f \in \mathcal{D}(L^*)$ satisfying $L^* f = \lambda f$ belongs to $\mathcal{A}$.

Notice also that in this more general setting, the Bakry-Émery curvature condition is the same as that in Definition C.8, where the sufficiently rich class corresponds to the extended algebra $\mathcal{A}$.