Characterizing (non-)Markovianity through Fisher Information

Paolo Abiuso*, Matteo Scandi, Dario De Santis and Jacopo Surace

ICFO-Institut de Ciencies Fotoniques, The Barcelona Institute of
Science and Technology, Castelldefels (Barcelona), 08860, Spain
* paolo.abiuso@icfo.eu

July 14, 2022

Abstract

A non-isolated physical system typically loses information to its environment, and when such loss is irreversible the evolution is said to be Markovian. Non-Markovian effects are studied by monitoring how information quantifiers, such as the distance between physical states, evolve in time. Here we show that the Fisher information metric emerges as the natural object to study in this context; we fully characterize the relation between its contractivity properties and Markovianity, both from the mathematical and operational point of view. We prove, for classical dynamics, that Markovianity is equivalent to the monotonous contraction of the Fisher metric at all points of the set of states. At the same time, operational witnesses of non-Markovianity based on the dilation of the Fisher distance cannot, in general, detect all non-Markovian evolutions, unless specific physical postprocessing is applied to the dynamics. Finally, we show for the first time that non-Markovian dilations of Fisher distance between states at any time correspond to backflow of information about the initial state of the dynamics at time 0, via Bayesian retrodiction. All the presented results can be lifted to the case of quantum dynamics by considering the standard CP-divisibility framework.

Contents

1   Introduction

The information contained in an evolving open system may undergo two possible dynamical regimes, dubbed Markovian or non-Markovian. Markovian evolutions are characterized by memoryless environments, where at each time the dynamical trajectory can be represented by physical transformations that solely depend on the immediately previous time step [1, 2]. It follows that in this regime the information contained in the system undergoes a monotonic degradation. On the other hand, non-Markovian dynamics are distinguished by memory effects, meaning that at any time the evolution might in general depend on all the previous time steps of the trajectory. This property allows information to flow back into the system. The complete characterization of non-Markovianity passes through the classification of the possible backflow phenomena that these evolutions can offer. This task is at the core of the non-Markovian witnessing problem [3, 4]. More specifically, one needs to find information quantifiers and specific initializations of the system (which may include ancillas) that are able to signal the non-Markovian nature of the evolution via revivals of an otherwise monotonically decaying information. Different quantifiers have been considered in this context, such as for state discrimination [5, 6], channel capacity [7], volume of accessible states [8] and correlations [9, 10]. This approach allows to understands how we can gain benefits from non-Markovian evolutions and, at the same time, it is committed to test the correspondence between its phenomenological and mathematical definition.

A quantity that has attracted a lot of attention from the community is the trace distance. In the classical case, it is given by

$$D_{\text{Tr}}(\boldsymbol{p}, \boldsymbol{q}) = |\boldsymbol{p} - \boldsymbol{q}| = \sum_i |p_i - q_i|, \tag{1}$$

where $\boldsymbol{p}$ and $\boldsymbol{q}$ are two classical probability distributions. Such distance quantifies the distinguishability between the states $\boldsymbol{p}$ and $\boldsymbol{q}$, as the error probability of distinguishing the two in a single-shot measurement is lower bounded by $P_{\text{err}} = \frac{2 - |\boldsymbol{p} - \boldsymbol{q}|}{4}$ [11]. Moreover, this quantity decreases under physical maps, leading to a monotonic decrease for Markovian evolutions [3] (cf. Appendix A). Hence, one consequence associated to the loss of information due to Markovianity is a continuous decrease in the ability of an agent to discriminate between any two states.

Still, the effects linked to Markovianity go well beyond what can be quantified by the trace distance alone. In the following we study an alternative quantifier: the Fisher information distance. This can be defined from its value between two infinitesimally close points. That is, given a small perturbation $|\boldsymbol{d}| \ll 1$ it reads

$$D_{\text{Fish}}^2(\boldsymbol{p}, \boldsymbol{p} + \boldsymbol{d}) \simeq \langle \boldsymbol{d}, \boldsymbol{d} \rangle_{\boldsymbol{p}} := \frac{1}{2} \sum_i \frac{d_i^2}{p_i}, \tag{2}$$

where we define the Fisher scalar product $\langle \boldsymbol{a}, \boldsymbol{b} \rangle_{\boldsymbol{p}} := \sum (a_i b_i)/(2p_i)$ [1]. This quantity has numerous operational interpretations: in metrology it is used to derive the Cramér-Rao bound [13, 14], a fundamental limit on the precision with which a parameter can be estimated; it quantifies the asymptotic distinguishability between multiple copies of two states (Chernoff bound [15]);

---

[1]Integrating Eq. (2) over the space of states one can obtain the general expression of the distance $D_{\text{Fish}}(\boldsymbol{p}, \boldsymbol{q}) = \sqrt{2} \arccos(\sqrt{\boldsymbol{p}} \cdot \sqrt{\boldsymbol{q}})$ [12].

moreover, it also arises as the infinitesimal expansion of the relative entropy [16] (in fact, it can be shown that any f-divergence locally behaves as the Fisher information [17]).

While the relation between Fisher metric and Markovianity has been previously partially analyzed [18] here we characterize it completely. In fact, a key characterisation of the Fisher metric is given by the Chentsov's theorem: this says that the Fisher information is the unique Riemannian metric that contracts under the action of all stochastic maps [19,20] (for the quantum case, see Petz [21]). This is the starting point of our work. In particular, we study whether this strong relation between stochasticity and contractivity of the Fisher information can be reversed. That is, is it true that a map is Markovian *if and only if* it contracts monotonically the Fisher information? What are the operational consequences of the contraction/dilation of such metric?

2   Framework

Before passing to the main treatment, we introduce here the main objects of interest for the rest of the work. A stochastic map (or *channel*) $T^{(t,0)}$ is a linear operator that evolves probability vectors (or *states*) $\boldsymbol{p} \in \mathcal{S}(\mathbb{R}^N)$, in the simplex defined as $\mathcal{S}(\mathbb{R}^N) := \{\boldsymbol{p} \in \mathbb{R}^N | p_i \geq 0 \wedge \sum_i p_i = 1\}$, from a time 0 to t,

$$T^{(t,0)}[\boldsymbol{p}(0)] = \boldsymbol{p}(t) \,. \tag{3}$$

It follows from this definition that, for every t,

$$\sum_i T_{ij}^{(t,0)} = 1 \,, \quad T_{ij}^{(t,0)} \geq 0 \ \forall \, i,j \,. \tag{4}$$

In fact, the elements $T_{ij}^{(t,0)}$ can be interpreted as the conditional probability of ending up in microstate i at time t starting from j at time 0, i.e. $T_{ij}^{(t,0)} = P(i,t|j,0)$.

Assuming the channel to be continuous and differentiable in time, as well as invertible [2], one can define the intermediate channel $T^{(t,s)}$ between two increasing times s and t as

$$T^{(t,s)} \equiv T^{(t,0)} \circ T^{(s,0)-1} \,, \quad t \geq s \geq 0 \,. \tag{5}$$

Channels with these properties constitutes the class of *smooth evolutions*. From its definition, it is easy to verify that $T^{(t,s)}$ still satisfies $\sum_i T_{ij}^{(t,s)} = 1 \ \forall j$, which can be equivalently rewritten as

$$T_{jj}^{(t,s)} = 1 - \sum_{i \neq j} T_{ij}^{(t,s)} \quad \forall j \,, \tag{6}$$

and corresponds to the requirement that the dynamics preserves the normalisation.

In the limit of infinitesimal time-steps, a smooth evolution $T^{(t,0)}$ is generated by the rate matrix

$$R^{(t)} \equiv \lim_{\delta t \to 0} \frac{T^{(t+\delta t,t)} - \mathbb{1}}{\delta t} \,, \tag{7}$$

such that $\frac{\mathrm{d}}{\mathrm{d}t} T^{(t,0)} = R^{(t)} \circ T^{(t,0)}$. Thanks to the condition in Eq. (6) (see also Eq. (10a) below), one can always decompose $R^{(t)}$ as

$$R^{(t)} = \sum_{i \neq j} a_{i \leftarrow j}^{(t)} (|i\rangle\langle j| - |j\rangle\langle j|) \,, \tag{8}$$

---

[2]Invertibility of the maps can can be always physical satisfied by introducing undetectable ε-noises to the dynamics.

where $a_{i \leftarrow j}^{(t)}$ are real coefficients called *rates*.

A smooth evolution T is called *stochastic-divisible* or *Markovian* [3] [2,3] if for any partition of the interval $[0, t]$ it can be split into intermediate channels, all of which are stochastic

$$T^{(t,0)} = T^{(t,t_{k-1})} \circ T^{(t_{k-1},t_{k-2})} \circ \cdots \circ T^{(t_1,0)} . \tag{9}$$

Since the composition of two Markovian evolutions is again Markovian, one can check the stochastic-divisibility of a channel by studying maps of the form $T^{(t+\delta t,t)} \approx \mathbb{1} + \delta t \, R^{(t)}$ for all possible times. is, we require $T^{(t+\delta t,t)}$ to be stochastic, imposing on the rate matrix, via Eq. (4),

$$\sum_i (\delta_{ij} + \delta t R_{i,j}^{(t)}) = 1 + \delta t \sum_i R_{i,j}^{(t)} = 1 \qquad \forall j, \tag{10a}$$

$$\delta_{ij} + \delta t R_{i,j}^{(t)} \geq 0 \qquad \forall \, i,j \,, \tag{10b}$$

where δ_{ij} denotes the Kronecker delta. The first conditions implies that $\sum_i R_{i,j}^{(t)} = 0$, which leads to the canonical decomposition (8). If we now focus on the second condition in Eq. (10) we see that $R_{i,j}^{(t)} \geq 0$ whenever $i \neq j$. In the parametrisation above this means that $a_{i \leftarrow j}^{(t)} \geq 0$. It follows that the rate matrices generating stochastic evolutions form a cone, whose elements are matrices of the form (8) having all the rates $a_{i \leftarrow j}$ positive. Since a smooth evolution T is Markovian if and only if $T^{(t+\delta t,t)}$ is stochastic $\forall t$, we can give the following definition:

Definition 1. A smooth evolution is Markovian if and only if for all times t the rates $a_{i \leftarrow j}^{(t)}$ are positive for all pairs of microstates $i \neq j$.

Remark. In the case of quantum dynamics the evolutions are given by Completely Positive and Trace Preserving (CPTP) superoperator $\mathcal{T}^{(t,0)}$ acting on density matrices, giving rise to the evolution $\mathcal{T}^{(t,0)}[\rho(0)] = \rho(t)$. In analogy to Eq. (9), we adopt the canonical definition of quantum Markovianity as CP-divisibility [3], i.e., $\mathcal{T}^{(t,0)}$ is said to be Markovian if for any partition of the interval $[0, t]$, it can be written as the composition of intermediate maps that are CPTP. As for classical dynamics, Markovianity of smooth quantum evolutions becomes equivalent to the positivity of the rates of the master equation [22]. For simplicity of exposition, all the following results are presented in the classical scenario, but thanks to a technical Lemma (Appendix C) we can lift them to the quantum regime. The precise proofs are deferred to Appendix D.

3 Contractivity of the Fisher metric and non-Markovianity detection

From now on we will omit time-dependency when no confusion can arise. The Fisher distance between any two points decreases under the action of stochastic maps. This directly implies that under Markovian dynamics the Fisher metric contracts continuously. Indeed, simple algebra (cf. Appendix B) yields, for infinitesimal $\boldsymbol{d}$,

$$\frac{\mathrm{d}}{\mathrm{d}t} D_{\mathrm{Fish}}^2(\boldsymbol{p}, \boldsymbol{p} + \boldsymbol{d}) = -\sum_{i \neq j} a_{i \leftarrow j} I_{i \leftarrow j} \leq 0 \,, \tag{11}$$

---

[3]Markovian evolutions are an essential tool for the description of multi-time process $\{t_0, t_1, \dots\}$ where the evolution of the state only depends on the latest previous sampling of it, i.e. $P(i_k, t_k | j_{k-1}, t_{k-1}; j_{k-2}, t_{k-2}; \dots) = P(i_k, t_k | j_{k-1}, t_{k-1})$, a condition equivalent to Eq. (9) [3].

where we implicitly defined

$$I_{i\leftarrow j} := \frac{1}{2}\left(\frac{d_i}{p_i} - \frac{d_j}{p_j}\right)^2 p_j \tag{12}$$

as the *Fisher information flow* associated to the rate $a_{i\leftarrow j}$. These are positive objects, so that the contraction of the Fisher metric is directly associated to the positivity of the rates $a_{i\leftarrow j}$.

Now suppose that for some time t there is a negative rate $a_{\tilde{i}\leftarrow\tilde{j}}^{(t)} < 0$ (i.e., the evolution is non-Markovian). Is this sufficient to reverse the contraction of the Fisher information? The positive answer is given by the following

Theorem 1. *A smooth evolution $T^{(t,0)}$ is Markovian if and only if it induces a decrease in Fisher distance between any two points in $\mathcal{S}(\mathbb{R}^N)$ at all times, i.e.*

$$T^{(t,0)} \text{ is Markovian } \Longleftrightarrow D_{\mathrm{Fish}}(T^{(t+\delta t,t)}[\boldsymbol{p}], T^{(t+\delta t,t)}[\boldsymbol{q}]) \le D_{\mathrm{Fish}}(\boldsymbol{p},\boldsymbol{q}) \,\forall t, \,\forall \boldsymbol{p},\boldsymbol{q} \in \mathcal{S}(\mathbb{R}^N). \tag{13}$$

Note that applying an intermediate map $T^{(t,s)}$ to all the points in the state space $\mathcal{S}(\mathbb{R}^N)$ is in general not physically defined, as only points in the image of $T^{(s,0)}$ are guaranteed to be physical states after applying $T^{(t,s)}$ (cf. Eq. (5)). Still, by considering the infinitesimal evolution $T^{(t+\delta t,t)} \simeq \mathbb{1} + \delta t R^{(t)}$, each point *in the interior* of $\mathcal{S}(\mathbb{R}^N)$ is guaranteed to be a physical state after evolving for a δt small enough. This allows $\frac{d}{dt} D_{\mathrm{Fish}}(\boldsymbol{p},\boldsymbol{q})|_t$ to be well defined at all times for all points in the interior of $\mathcal{S}(\mathbb{R}^N)$.

Proof. As the $\Rightarrow$ implication in Eq. (13) is trivial from the above discussion, only the proof of the opposite $\Leftarrow$ is needed. For that, suppose that the first instance of non-Markovianity happens between time t and $t + \delta t$. In order to prove the statement it is sufficient to consider the Fisher distance between any two infinitesimally close points $\boldsymbol{p}$ and $\boldsymbol{p} + \boldsymbol{d}$, with $|\boldsymbol{d}| \ll 1$, evolving according to $T^{(t+\delta t,t)}$. Then, for any negative rate $a_{\tilde{i}\leftarrow\tilde{j}}^{(t)} < 0$, one can find a point $\boldsymbol{p}$ and a perturbation $\boldsymbol{d}$ such that $\frac{d}{dt} D_{\mathrm{Fish}}^2(\boldsymbol{p},\boldsymbol{p}+\boldsymbol{d}) > 0$. In fact, assume without loss of generality that $a_{1\leftarrow 2} < 0$ and consider $\boldsymbol{p}$ and $\boldsymbol{d}$ of the form

$$\boldsymbol{p} = \begin{pmatrix} \mathcal{O}(\varepsilon) \\ 1-\varepsilon \\ \mathcal{O}(\varepsilon) \\ \vdots \\ \mathcal{O}(\varepsilon), \end{pmatrix}, \quad \boldsymbol{d} = \begin{pmatrix} \mathcal{O}(\varepsilon) \\ \mathcal{O}(\varepsilon) \\ \mathcal{O}(\varepsilon^2) \\ \vdots \\ \mathcal{O}(\varepsilon^2) \end{pmatrix}, \tag{14}$$

where ε is an arbitrary small number and we assume that the vectors $\boldsymbol{p}$ and $\boldsymbol{p} + \boldsymbol{d}$ are properly normalised. Notice that it is always possible to choose $\boldsymbol{p}$ and $\boldsymbol{p} + \boldsymbol{d}$ in the interior of $\mathcal{S}(\mathbb{R}^N)$ (i.e., with strictly positive components). Inserting this expression in Eq. (11) we find that the only term of order $\mathcal{O}(1)$ comes from setting $i = 1, j = 2$ in the sum above. That is, at leading order,

$$\frac{d}{dt} D_{\mathrm{Fish}}^2(\boldsymbol{p},\boldsymbol{p}+\boldsymbol{d}) = -a_{1\leftarrow 2} \frac{d_1^2}{p_1^2} + \mathcal{O}(\varepsilon). \tag{15}$$

Since $a_{1\leftarrow 2} < 0$ we can always find a ε small enough so that this quantity is strictly positive. Hence, for any non-Markovian dynamics there always exists $\boldsymbol{p}$ and $\boldsymbol{d}$ such that $D_{\mathrm{Fish}}^2(\boldsymbol{p},\boldsymbol{p}+\boldsymbol{d})$ locally increases, proving the Theorem. $\square$

Theorem 1 can be seen as a completion of Chentsov's Theorem, as it implies that not only the Fisher information decreases under Markovian evolutions, but also that an evolution contracting the Fisher distance between any two points has to be Markovian. The proof generalizes to the case of quantum dynamics in the canonical CP-divisibility framework [3], by considering a copy of the system on which the dynamics acts trivially, i.e., the evolution is given by $\mathcal{T}^{(t,0)} \otimes \mathbb{1}_N$ and the set of states is considered on the global bipartition (cf. Appendix D.1).

Interestingly, a similar theorem cannot hold for the trace-distance as one can explicitly construct non-Markovian evolutions that monotonically contract $D_{\mathrm{Tr}}(\boldsymbol{p}, \boldsymbol{q})$ for any two points (see Appendix A). This corroborates the interpretation of the Fisher metric as the canonical distance whose contractivity identifies stochastic-divisible maps. Still, the argument that lead to Thm. 1 has a shortcoming: even if non-Markovianity implies the dilation of the Fisher information, this is not sufficient to produce an operational witness. In fact, assume that at time t there is a local dilation for two points close to $\tilde{\boldsymbol{p}}$. In order to observe it, one would need to initialise the system in the state $\boldsymbol{p}(0) = (T^{(t,0)})^{-1} \tilde{\boldsymbol{p}}$. If $\tilde{\boldsymbol{p}}$ is outside the image of $T^{(t,0)}$, this cannot be achieved physically. That is, the drawback of this approach is that the Fisher metric is point-dependent, and the witnessing point might be excluded by the dynamics $T^{(t,0)}$. On the other hand, the trace-distance is translational invariant, i.e., $D_{\mathrm{Tr}}(\boldsymbol{p}, \boldsymbol{q}) = |\boldsymbol{d}|$ where $\boldsymbol{d} = \boldsymbol{p} - \boldsymbol{q}$. Then, as soon as any two points $\boldsymbol{p}$ and $\boldsymbol{q}$ are increasing their trace-distance, in order to present an operational witness it is sufficient to consider a point $\boldsymbol{r}$ in the interior of the image of $T^{(t,0)}$ and ε small enough for $\boldsymbol{r}_\varepsilon = \boldsymbol{r} + \varepsilon(\boldsymbol{p} - \boldsymbol{q})$ to be in the image as well. Then, $D_{\mathrm{Tr}}(\boldsymbol{r}, \boldsymbol{r}_\varepsilon) = \varepsilon D_{\mathrm{Tr}}(\boldsymbol{p}, \boldsymbol{q})$ and this increases by assumption. Moreover, if one adds a finite number of ancillary degrees of freedom on which the dynamics acts trivially, one can always find such two points for any non-Markovian evolution (see [5] and App A). A similar property does not hold for the Fisher distance, as it lacks translation invariance. More specifically,

Theorem 2. *No finite number n of copies of the channel $T^{(t,0)}$ nor ancillary degrees of freedom of any dimension M is enough to witness all non-Markovian evolutions via revivals of the Fisher distance between two initially prepared states.*

Specifically, given n copies of the system, and an ancilla with arbitrary dimension M, the state space will be $\mathcal{S}(\mathbb{R}^{N^{\otimes n}} \otimes \mathbb{R}^M)$ and the dynamics acting on it $\bar{T}^{(t,0)} = T^{(t,0)\otimes n} \otimes \mathbb{1}_M$.

Proof. We constructively provide, for any n and M, a counterexample in which all the states in the image of $\bar{T}^{(t,0)}$ continue decreasing their Fisher distance between time t and $t + \delta t$, even if $T^{(t+\delta t, t)} \simeq \mathbb{1} + \delta t R^{(t)}$ is non-stochastic. Here we provide the proof for the single copy case, deferring the multiple cases one to Appendix D.2. The map we consider is then given by $\bar{T}^{(t,0)} = T^{(t,0)} \otimes \mathbb{1}_M$ and the rate matrix by $\bar{R}^{(t)} = \frac{\mathrm{d}}{\mathrm{d}t} \bar{T}^{(t,0)} = R^{(t)} \otimes \mathbb{1}_M$. Suppose now that there is a unique negative rate $a_{\tilde{i} \leftarrow \tilde{j}}$ and that the image of $T^{(t,0)}$ is contained to a small ball around an appropriate vector $\boldsymbol{\pi}$ (e.g., by a map of the form $T^{(t,0)}[\boldsymbol{p}] = \boldsymbol{\pi}(1-\varepsilon) + \varepsilon \boldsymbol{p}$). Attach at time 0 an arbitrary ancilla, so that the initial state is given by $\boldsymbol{p}(0) \in \mathcal{S}(\mathbb{R}^N \otimes \mathbb{R}^M)$, and define $\boldsymbol{w}$ to be its reduced marginal on $\mathbb{R}^M$, whereas the dynamics is given by $\bar{T}^{(t,0)} = T^{(t,0)} \otimes \mathbb{1}_M$. Then, the state at time t will be ε-close to

$$\boldsymbol{p}(t) \sim \boldsymbol{\pi} \otimes \boldsymbol{w} + \mathcal{O}(\varepsilon). \tag{16}$$

Notice also that the rate matrix $\bar{R}^{(t)} = \frac{\mathrm{d}}{\mathrm{d}t} \bar{T}^{(t,0)} = R^{(t)} \otimes \mathbb{1}_M$ has the following coordinate expression

$$[\bar{R}]_{ij,\alpha\beta} = R_{ij}\delta_{\alpha\beta} \quad i,j \in \{1,\ldots,N\} \quad \alpha,\beta \in \{1,\ldots,M\}, \tag{17}$$

so that the rates are simply given by $a_{i\alpha \leftarrow j\beta} = a_{i\leftarrow j}\delta_{\alpha\beta}$. In this scenario, the evolution of the

Fisher distance (as expressed in Eq. (11)) becomes

$$-\sum_{i\neq j,\alpha} a_{i\leftarrow j} \left(\frac{d_{i\alpha}}{p_{i\alpha}} - \frac{d_{j\alpha}}{p_{j\alpha}}\right)^2 p_{j\alpha} + \mathcal{O}(\varepsilon) \quad \text{with } p_{j\alpha} = \pi_j w_\alpha. \tag{18}$$

Again, consider the case in which at time t a single rate becomes negative, for definiteness say $a_{1\leftarrow 2} < 0$. Then, it is sufficient that $a_{2\leftarrow 1}\pi_1 > |a_{1\leftarrow 2}|\pi_2$ to see that the sum in Eq. (18) is strictly negative in the limit $\varepsilon \to 0$. Hence, even if the dynamics is non-Markovian, there is no increase in Fisher distance on the image of $T^{(t,0)}$, proving Theorem 2 for $n = 1$. $\square$

In the same way, even using multiple copies of the channel does not help finding a witness. The proof easily generalises to $n \geq 2$, as presented in App. D.2. It should be noticed that the condition of finite copies in Thm. 2 cannot be dropped: in fact, in the limit of infinite copies, one can perform full tomography of the evolution, allowing to reconstruct its action also on points outside of the image of $T^{(t,0)}$.

Despite the above "no-go" Theorem, we can introduce an operational non-Markovianity witness which does not require additional copies of the channel, but only some specific post-processing of the states. More specifically, the following technical theorem holds

Theorem 3. *For any state p and perturbation d on $\mathcal{S}(\mathbb{R}^N \otimes \mathbb{R}^M)$ (where we admit an M-dimensional ancilla), it is possible to implement a class of transformations F_d depending on d and on $T^{(t,0)}$ that witness non-Markovianity at time t. That is, if $T^{(t+\delta t,t)}$ is stochastic, for any choice of d, one has that*

$$D_{\text{Fish}}(F_d \circ T^{(t+\delta t,0)}[p], F_d \circ T^{(t+\delta t,0)}[p+d]) \leq D_{\text{Fish}}(F_d \circ T^{(t,0)}[p], F_d \circ T^{(t,0)}[p+d]) \tag{19}$$

whereas in the presence of non-Markovianity (i.e., for $T^{(t+\delta t,t)}$ not stochastic) there exists at least one d for which the inequality is reversed (i.e., the Fisher distance of the post-processed states increases). Moreover, for classical systems $M = 2$ is enough to witness in this way all non-Markovian evolutions.

The specific proof is given in Appendix D.3. Theorem 3 ensures that any break of stochastic-divisibility in the interval $[t, t+\delta t]$ can be operationally witnessed via backflow of Fisher information between states that undergo the transformation F_d before being measured. The specific construction that we used here requires the knowledge of the previous dynamics $T^{(t,0)}$, which makes such witnessing unpractical. Still, this protocol should be considered as a proof of principle of the possibility of witnessing non-Markovianity through post-processing.

4 Backflows of information from Bayesian retrodiction

In most of the literature about non-Markovianity backflows of information are considered by studying states $p(t)$ at time t [3, 4], while the question about backflows of information *about the initial state* $p(0)$ remains instead largely unexplored. Even if one can argue that the in-vertibility of the dynamics preserves the information about the initial conditions (as these can be recovered as $p(0) = T^{(t,0)^{-1}}[p(t)]$), actually retrieving the initial state from $p(t)$ requires full tomography both of the state and of the channel, and a post-processing of such data which cannot be performed physically in a single-shot scenario.

Conversely, we consider here a Bayesian inversion of $p(t)$ that allows us to compare it with the initial state through a physically implementable transformation. In particular, define the

prior as a vector π representing our knowledge about the system at time 0. Given an evolution $T^{(t,0)}$, the Bayes-recovery map is defined as

$$\hat{T}_t = J_\pi \circ T_t^{\mathsf{T}} \circ J_{T[\pi]}^{-1} \tag{20}$$

where we use the shorthand notation $T_t := T^{(t,0)}$, and we introduced the map J_p, corresponding to the diagonal operator that multiplies each component of a vector by the corresponding component of p, i.e. $[J_p]_{ij} = \delta_{ij} p_j$.

The map $\hat{T}_t$ is stochastic (i.e., physically implementable) and represents a recovery of the state via statistical retrodiction [23, 24]. Indeed if one identifies the coordinates of the prior with the corresponding probability, i.e., $\pi_i \equiv P(i, 0)$, and the components of the maps with the corresponding transitions, i.e., $(T_t)_{ij} \equiv P(i, t | j, 0)$, it is easy to verify that $(\hat{T}_t)_{ij} \equiv P(i, 0 | j, t)$ satisfies the Bayes rule. It should also be noticed that $\hat{T}_t$ perfectly recovers the prior at all times ($\pi = \hat{T}_t \circ T_t[\pi]$).

This channel allows us to study how much information is stored in the evolved state $p(t)$ about its initial conditions. In particular, we can compare the distance between $p(0)$ and the retrodicted state $\hat{p}(t) := \hat{T}_t[p(t)] = \hat{T}_t \circ T_t[p(0)]$. We also assume that the prior contains some knowledge on the initial conditions, so that we can write $p(0) = \pi + d$, for some small $|d| \ll 1$. Moreover, since π is perfectly recovered, we also have that $\hat{p}(t) = \pi + \hat{d}(t)$. Then, the Fisher information between $p(0)$ and $\hat{p}(t)$ reads

$$D_{\mathrm{Fish}}^2(p(0), \hat{p}(t)) \simeq \left\langle d - \hat{d}(t), d - \hat{d}(t) \right\rangle_\pi . \tag{21}$$

Interestingly, this object is directly connected to the Fisher information at time t:

Theorem 4. *The contractivity of the Fisher information at time t is in one-to-one correspondence with the expansivity of Eq.* (21). *That is,*

$$\frac{\mathrm{d}}{\mathrm{d}t} D_{\mathrm{Fish}}^2(p(0), \hat{p}(t)) = \frac{\mathrm{d}}{\mathrm{d}t} \left\langle d - \hat{d}(t), d - \hat{d}(t) \right\rangle_\pi \geq 0 \tag{22}$$

if and only if the Fisher information contracts for any two points in the vicinity of $T_t[\pi]$.

Proof. The main ingredient in the proof of this theorem is given by the following identities

$$\langle d, \hat{T}_t T_t[d] \rangle_\pi = \langle T_t[d], T_t[d] \rangle_{T_t[\pi]} = \langle \hat{T}_t T_t[d], d \rangle_\pi, \tag{23}$$

which can verified by directly substituting $\hat{T}_t$ with its definition in Eq. (20). One can read from these equalities the following two facts: first, $\hat{T}_t$ can be used to put in relation the Fisher information at time 0 and at time t; secondly, $\hat{T}_t T_t$ is self-adjoint with respect to $\langle \bullet, \bullet \rangle_\pi$. This allows to rewrite Eq. (22) as

$$\frac{\mathrm{d}}{\mathrm{d}t} \left\langle d, (\mathbb{1} - \hat{T}_t T_t)^2[d] \right\rangle_\pi = -2 \left\langle d, (\mathbb{1} - \hat{T}_t T_t) \frac{\mathrm{d}}{\mathrm{d}t} \hat{T}_t T_t[d] \right\rangle_\pi . \tag{24}$$

First notice that $(\mathbb{1} - \hat{T}_t T_t)$ is positive definite. In fact, this can be seen from

$$\langle d, \hat{T}_t T_t[d] \rangle_\pi = \langle T_t[d], T_t[d] \rangle_{T_t[\pi]} \leq \langle d, d \rangle_\pi , \tag{25}$$

where the last inequality follows from the contractivity of the Fisher information. Moreover, we also have that $\langle d, \frac{\mathrm{d}}{\mathrm{d}t} \hat{T}_t T_t[d] \rangle_\pi = \frac{\mathrm{d}}{\mathrm{d}t} \langle T_t[d], T_t[d] \rangle_{T_t[\pi]}$, so that $-\frac{\mathrm{d}}{\mathrm{d}t} \hat{T}_t T_t$ is positive if and only if the Fisher contracts monotonically. If this is the case, we can use the fact that the product of two positive operators has positive spectrum. On the other hand, if the Fisher metric is expanding at time t, there exists an eigenvector $\tilde{d}$ of $-\frac{\mathrm{d}}{\mathrm{d}t} \hat{T}_t T_t$ with negative eigenvalue $\lambda < 0$, so that

$$-2 \left\langle \tilde{d}, (\mathbb{1} - \hat{T} T) \frac{\mathrm{d}}{\mathrm{d}t} \hat{T} T[\tilde{d}] \right\rangle_\pi = 2\lambda \left\langle \tilde{d}, (\mathbb{1} - \hat{T} T)[\tilde{d}] \right\rangle_\pi < 0, \tag{26}$$

proving the claim. □
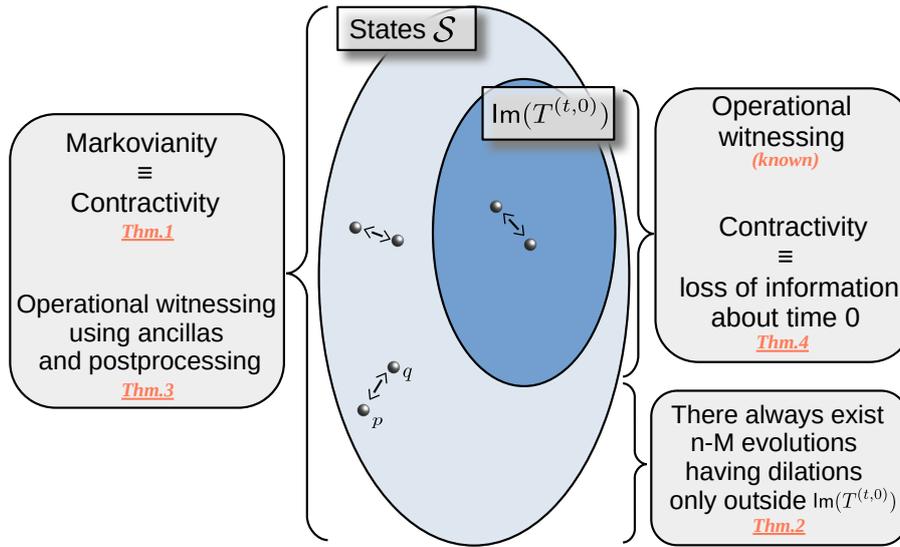
Figure 1: The main quantity analyzed in this work is the rate of contraction/dilation of the Fisher metric, i.e., $\frac{d}{dt}D_{\text{Fish}}(\boldsymbol{p},\boldsymbol{q})$, with $|\boldsymbol{p}-\boldsymbol{q}| \ll 1$, when both $\boldsymbol{p}$ and $\boldsymbol{q}$ evolve according to the local intermediate map $T^{(t+\delta t,t)}$. We characterize the mathematical and operational meaning of the negativity/positivity of such rate, both inside and outside the image of the evolution $T^{(t,0)}$.

This theorem tells us that the ability of an agent of retrieving the initial state of the dynamics decreases under Markovian evolution. Moreover, in the case in which there are backflows in the Fisher information, non-Markovianity helps obtaining more information about the initial state.

5 Conclusions

In this work we characterized the relation between Markovianity, Fisher metric contractivity and information flow, both from the mathematical and operational point of view. The resulting picture can be seen in Fig. 1: we showed that monotonous contractivity of the Fisher metric on the whole set of states and at all times, is mathematically equivalent to Markovianity (Thm. 1). As known, when the metric dilates locally inside the image of the evolution $\text{Im}(T^{(t,0)})$, a backflow of Fisher information can be operationally witnessed. At the same time, non-Markovian evolutions might in general show Fisher metric dilations only outside the image of the evolution itself, regardless of the number of copies of the channel and ancillary degrees of freedom available (Thm. 2). To witness operationally non-Markovianity in such cases, one needs postprocessing to be appended to the dynamics (Thm. 3). Finally, we showed that dilations of the Fisher metric between evolving states can be mapped to a backflow of information about the initial states by applying Bayesian retrodiction (Thm. 4).

Our results corroborate the idea that the Fisher Information metric is the natural object whose contractivity properties characterize memory effects in open system dynamics, both from the mathematical and operational point of view.

5.1 The case of quantum dynamics

The results were presented for classical dynamics in order to keep the exposition clean. In fact, whereas the Fisher information distance is uniquely defined for classical systems, the

presence of non commutative observables allows only for the definition of a family of Fisher information in the quantum case (see Appendix C). Despite this technical difficulty, we also show in Appendix C that in order to study the interplay between Markovianity and Fisher information we can limit ourselves to the diagonal subspaces, practically going back to the classical constructions (details for each case are given in Appendix D). The only case in which we cannot ignore the non-uniqueness of the quantum Fisher information are the expression of the Fisher information flows Eq. (12), and Thm. 4. Both of these results will be generalised to the quantum regime in the forthcoming work [25].

Moreover, it should be noticed that for quantum dynamics we identify Markovianity with CP-divisibility [3]. This enforces an important difference with the classical case: whereas for the classical dynamics the minimal space that can be used to probe the Markovianity of an N-dimensional system is given by $\mathcal{S}(\mathbb{R}^N)$, the corresponding for quantum systems is given by two copies of the same Hilbert space on which the dynamics acts as $\mathcal{T}^{(t,0)} \otimes \mathbb{1}_N$. In fact, it is not possible to distinguish P-divisibility from CP-divisibility by restricting to the system space alone without the use of any ancillas.

A   The trace distance: contractivity and witnesses

The trace distance between two classical states is given by $|\boldsymbol{p} - \boldsymbol{q}|_{\mathrm{Tr}} \equiv \sum |p_i - q_i|$. Interestingly, it only depends on the vector $\boldsymbol{d} := \boldsymbol{q} - \boldsymbol{p}$ and not on the base-point $\boldsymbol{p}$. Distances satisfying this condition are called *translational invariant*. Then, we can write its evolution as:

$$\frac{\mathrm{d}}{\mathrm{d}t} D_{\mathrm{Tr}}(\boldsymbol{p}, \boldsymbol{q}) = \frac{\mathrm{d}}{\mathrm{d}t} |\boldsymbol{d}|_{\mathrm{Tr}} = \sum_i \frac{\mathrm{d}}{\mathrm{d}t} |d_i| = \sum_i \mathrm{sign}(d_i) \dot{d}_i. \tag{27}$$

For classical systems, given a rate matrix decomposed as in Eq. (8), one has that

$$\dot{d}_i = \sum_j R_{i,j} d_j = \sum_{\substack{j \\ j \neq i}} \left(a_{i \leftarrow j} d_j - a_{j \leftarrow i} d_i \right), \tag{28}$$

so that we can rewrite Eq. (27) as

$$\sum_{i,j,i \neq j} \mathrm{sign}(d_i) \left(a_{i \leftarrow j} d_j - a_{j \leftarrow i} d_i \right) = \sum_{i,j,i \neq j} \left(\mathrm{sign}(d_j) a_{j \leftarrow i} d_i - \mathrm{sign}(d_i) a_{j \leftarrow i} d_i \right) = \tag{29}$$

$$= \sum_{i,j,i \neq j} \left(\mathrm{sign}(d_j) - \mathrm{sign}(d_i) \right) a_{j \leftarrow i} d_i \leq 0, \tag{30}$$

where the last expression is clearly negative whenever $a_{j \leftarrow i}$ are positive: in fact, either d_j and d_i have equal sign, in which case the term is null, or $\mathrm{sign}(d_j) = -\mathrm{sign}(d_i)$, so that the factor $(\mathrm{sign}(d_j) - \mathrm{sign}(d_i)) = -2\,\mathrm{sign}(d_i)$, and we can rewrite the terms in the sum as $-2\,\mathrm{sign}(d_i)d_i = -2|d_i|$.

This shows explicitly how the trace distance decreases under Markovian maps. Is the reverse true? That is, if the evolution stops being stochastic-divisible, i.e., some $a_{\tilde{i} \leftarrow \tilde{j}}$ is negative,

can one always find two points for which the trace distance between them increases? Differently from the Fisher distance, the answer to this question is negative: one easy counterexample can be given in dimension $N = 2$, with $a_{1\leftarrow 2} < 0$ and $a_{2\leftarrow 1} > 0$ with $|a_{2\leftarrow 1}| > |a_{1\leftarrow 2}|$. Plugging such choice in Eq. (30), and using the fact that in dimension 2, $\boldsymbol{d}$ satisfies $d_1 = -d_2$, it is easy to check that the derivative of $|\boldsymbol{d}|_{\text{Tr}}$ stays negative.

This fact could sound counter-intuitive: it is well-known that a map is stochastic *if and only if* each vector $\boldsymbol{v}$ decreases in trace-norm [3]. Yet, we have proven here that the trace distance cannot witness all non-Markovian evolutions, without resorting to ancillas. An easy resolution of this paradox is given by noticing that the test vectors used here are all in the form $\boldsymbol{d} = \boldsymbol{p} - \boldsymbol{q}$, constraining them to be of zero trace. This lowers the dimension of the tested vectors by one, resolving the contradiction. It is in fact straightforward to verify that simply by choosing $v_i = \delta_{i,2}$ in our example above one would be able to witness non-Markovianity (where $\delta_{i,j}$ is the Kronecker delta). Still, the trace condition prevents us from accessing this vector.

If one allows for the use of ancillas, though, the situation changes. In fact, consider an ancillary set $\mathcal{S}(\mathbb{R}^M)$ on which the dynamics acts trivially (i.e., states belong to $\mathcal{S}(\mathbb{R}^N \otimes \mathbb{R}^M)$ and the dynamics is of the form $\bar{T}^{(t,0)} = T^{(t,0)} \otimes \mathbb{1}_M$). It is now sufficient to consider the following vector on the extended space

$$\boldsymbol{d} = \boldsymbol{v} \otimes \boldsymbol{d}_{\text{anc}}, \quad \boldsymbol{v} \in \mathbb{R}^N, \ v_i = \delta_{i,2}, \quad \boldsymbol{d}_{\text{anc}} \in \mathbb{R}^M, \quad \text{Tr}[\boldsymbol{d}_{\text{anc}}] = 0, \tag{31}$$

to witness the non-Markovianity in the example above. Notice that the trace condition on $\boldsymbol{d}_{\text{anc}}$ ensures that $\boldsymbol{d}$ is traceless, so this is a valid distance vector. Then, since its derivative takes the form $\dot{\boldsymbol{d}} = \dot{\boldsymbol{v}} \otimes \boldsymbol{d}_{\text{anc}}$, the trace distance increases in time

$$\frac{\mathrm{d}}{\mathrm{d}t}|\boldsymbol{d}| = |\dot{\boldsymbol{v}} \otimes \boldsymbol{d}_{\text{anc}}| = |\dot{\boldsymbol{v}}||\boldsymbol{d}_{\text{anc}}| > 0, \tag{32}$$

as $\boldsymbol{v}$ was chosen to have $|\dot{\boldsymbol{v}}| > 0$ when $a_{1\leftarrow 2} < 0$. Notice that the ancilla can have dimension as small as 2. The same kind of reasoning is used in Ref. [5] for the case of quantum dynamics. From these simple considerations and from reference [5] we have the following

Lemma 1. *Given a non-stochastic-divisible dynamics $T^{(t,0)}$, adding a finite ancilla and considering the dynamics $T^{(t,0)} \otimes \mathbb{1}_M$ allows for witnessing any such dynamics via revivals in trace distance between initially prepared states. A finite number M of ancillary degrees of freedom is enough, and in the classical case $M = 2$.*

In the quantum case, $M = N + 1$ is needed, where N is the dimension of the quantum state [5]. We close the section with a final remark. Notice that even *less* than a finite ancilla would be enough. By this we mean that it is enough to enlarge the state space with an additional microstate that does not interact with the others, i.e., considering $\mathcal{S}(\mathbb{R}^N) \to \mathcal{S}(\mathbb{R}^{N+1})$ and a dynamics of the form $T' = T \oplus 1$ (i.e., $T'_{ij} = T_{ij}$ for $i, j \in \{1, \dots, N\}$ and $T'_{ij} = \delta_{ij}$ if one between i and j is the $N + 1$ index). Given $\boldsymbol{v} \in \mathbb{R}^N$ such that $|\dot{\boldsymbol{v}}| > 0$, one can consider $\boldsymbol{d} \in \mathbb{R}^{N+1}$ with

$$\boldsymbol{d} = \begin{cases} d_i = v_i & \text{if } i \in \{1, \dots, N\} \\ d_i = -\sum_{i=1}^N v_i & \text{if } i \equiv N + 1 \end{cases} \tag{33}$$

Since the map T is trace-preserving at all times, such vector satisfies

$$|\dot{\boldsymbol{d}}| = |\dot{\boldsymbol{v}}|, \tag{34}$$

allowing for the witnessing of non-Markovianity. This result should be compared with the discussion above about how the trace condition lowers the accessible dimension by one.

B   Rates of Fisher Information flow and contractivity

We present here a decomposition of the derivative of the Fisher information in a sum of independent flows. In particular, by explicitly using the expression in Eq. (8) for the dynamics we can compute

$$
2\frac{\mathrm{d}}{\mathrm{d}t}D_{\mathrm{Fish}}^2(\boldsymbol{p}, \boldsymbol{p}+\boldsymbol{d}) = \sum_i \left(\frac{2d_i \dot{d}_i}{p_i} - \frac{d_i^2}{p_i^2}\dot{p}_i \right) =
$$

$$
= \sum_{i,j} \left(\frac{2d_i(a_{i\leftarrow j}d_j - a_{j\leftarrow i}d_i)}{p_i} - \frac{d_i^2}{p_i^2}(a_{i\leftarrow j}p_j - a_{j\leftarrow i}p_i) \right) =
$$

$$
= \sum_{i,j} \left(\frac{2d_i d_j a_{i\leftarrow j}}{p_i} - \frac{d_i^2}{p_i}a_{j\leftarrow i} - \frac{d_i^2}{p_i^2}p_j a_{i\leftarrow j} \right) =
$$

$$
= \sum_{i,j} a_{i\leftarrow j} \left(\frac{2d_i d_j}{p_i} - \frac{d_j^2}{p_j} - \frac{d_i^2}{p_i^2}p_j \right) =
$$

$$
= -\sum_{i\neq j} a_{i\leftarrow j} \left(\frac{d_i}{p_i} - \frac{d_j}{p_j} \right)^2 p_j \leq 0 \,, \quad (35)
$$

where between the third and the fourth line we made the change of variable $i \to j$ in order to factor out $a_{i\leftarrow j}$. Whenever the evolution is Markovian (i.e., the rates $a_{i\leftarrow j}$ are all positive), the Fisher information is contracting, in analogy with what happened for the trace distance. On the other hand, if one compares the result in Eq. (30) with what we just obtained, there is an important difference between the two: whereas the evolution of the trace distance only depends on $\boldsymbol{d}$ (mirroring the translational invariance of this quantity), the Fisher information explicitly depends on the base-point. This difference will be particularly important in the constructions of the next sections.

C   The quantum Fisher information metric and its relation to the trace distance

The extension of the Fisher information to quantum systems is done by generalising Chentsov's theorem to completely positive, trace preserving maps (CPTP). That is, a metric on quantum states is called monotone if it decreases under all CPTP maps. Then, it was shown by Petz in [21] that all such metrics are induced by scalar products of the form:

$$
K_\rho^f(A, B) := \frac{1}{2}\mathrm{Tr}\left[A^\dagger \, \mathbb{J}_{f,\rho}^{-1}[B] \right], \quad (36)
$$

where $\mathbb{J}_{f,\rho}$ is a self-adjoint superoperator given by:

$$
\mathbb{J}_{f,\rho} := \mathbb{R}_\rho \, f(\mathbb{L}_\rho \mathbb{R}_\rho^{-1}), \quad (37)
$$

and $\mathbb{L}_\rho/\mathbb{R}_\rho$ are the left/right multiplication operators acting as $\mathbb{L}_\rho \pi = \rho \pi$ (respectively $\mathbb{R}_\rho \pi = \pi\rho$) and $f : \mathbb{R}^+ \to \mathbb{R}^+$ is a standard operator monotone function. Despite its complicated form, the interpretation of $K_\rho^f(A, B)$ as the natural extension of the Fisher information to the quantum regime is corroborated by the result in [26], where it was shown that the same quantity emerges from the local expansion of the quantum generalisation of Csizár divergences. For this reason, we define the family of quantum Fisher distances as

$$
D_{\mathrm{Fish},f}^2(\rho, \rho + \delta\rho) \simeq K_\rho^f(\delta\rho, \delta\rho) := \frac{1}{2}\mathrm{Tr}\left[\delta\rho^\dagger \, \mathbb{J}_{f,\rho}^{-1}[\delta\rho] \right]. \quad (38)
$$

The uniqueness of the classical Fisher metric is hence substituted with a whole family of different monotone metrics. It is interesting to point out, though, that when $\mathbb{J}_{f,\rho}$ acts on diagonal states it behaves (independently of f) as the multiplication by ρ. That is, if $[A,\rho]=[B,\rho]=0$, with a small abuse of notation we have

$$K_\rho^f(A,B) := \frac{1}{2}\mathrm{Tr}\left[\frac{A^\dagger B}{\rho}\right] = \langle A,B\rangle_\rho\,, \tag{39}$$

that is all the quantum Fisher metrics collapse to the classical one for diagonal states. This feature allows us to lift most of our classical constructions to the quantum scenario without further complications.

Specifically, the following Lemma will be particularly relevant:

Lemma 2. *Given a state $\rho = \sum_i \rho_i\,|i\rangle\langle i|$, any perturbation $\delta\rho$ can be decomposed in diagonal and coherent part as:*

$$\delta\rho = \delta\rho_\Delta + \delta\rho_C \quad \text{with} \quad [\rho,\delta\rho_\Delta]=0 \quad \text{and} \quad \langle i|\delta\rho_C\,|i\rangle = 0\,, \tag{40}$$

that is, $\delta\rho_\Delta$ is the diagonal part of the matrix $\delta\rho$ and $\delta\rho_C$ the off-diagonal. Then for all f

$$\mathrm{Tr}\left[\delta\rho\,\mathbb{J}_{f,\rho}^{-1}[\delta\rho]\right] = \mathrm{Tr}\left[\delta\rho_\Delta\,\mathbb{J}_{f,\rho}^{-1}[\delta\rho_\Delta]\right] + \mathrm{Tr}\left[\delta\rho_C\,\mathbb{J}_{f,\rho}^{-1}[\delta\rho_C]\right] = \tag{41}$$

$$= 2\,\langle\boldsymbol{\delta},\boldsymbol{\delta}\rangle_{\boldsymbol{\rho}} + \mathrm{Tr}\left[\delta\rho_C\,\mathbb{J}_{f,\rho}^{-1}[\delta\rho_C]\right]\,, \tag{42}$$

where the components of the vectors $\boldsymbol{\delta}$, $\boldsymbol{\rho}$ are specified as

$$\boldsymbol{\rho}_i = \langle i|\rho\,|i\rangle\,, \quad \boldsymbol{\delta}_i = \langle i|\delta\rho\,|i\rangle = \langle i|\delta\rho_\Delta\,|i\rangle\,. \tag{43}$$

This result directly follows from the fact that

$$\mathrm{Tr}\left[\delta\rho_\Delta\,\mathbb{J}_{f,\rho}^{-1}[\delta\rho_C]\right] = \mathrm{Tr}\left[\delta\rho_C\,\mathbb{J}_{f,\rho}^{-1}[\delta\rho_\Delta]\right] = 0\,, \tag{44}$$

since $\mathbb{J}_{f,\rho}^{-1}[\delta\rho_\Delta]$ is itself diagonal in the basis $|i\rangle$ of eigenvectors of ρ. Then, we can use the Lemma above to prove the following corollary:

Corollary 1. *Consider a perturbation of the form $\delta\rho = \delta\rho_\Delta + \mathrm{d}t\,\delta\rho_C$, where $\mathrm{d}t$ is an infinitesimal quantity. Then, from Lemma 2 it follows that*

$$\mathrm{Tr}\left[\delta\rho\,\mathbb{J}_{f,\rho}^{-1}[\delta\rho]\right] = \mathrm{Tr}\left[\delta\rho_\Delta\,\mathbb{J}_{f,\rho}^{-1}[\delta\rho_\Delta]\right] + \mathcal{O}\left(\mathrm{d}t^2\right) = 2\,\langle\boldsymbol{\delta},\boldsymbol{\delta}\rangle_{\boldsymbol{\rho}} + \mathcal{O}\left(\mathrm{d}t^2\right). \tag{45}$$

In particular, the time derivative of the Fisher Information between ρ and $\rho+\delta\rho$ for $[\rho,\delta\rho]=0$ coincides with the derivative of the classical Fisher Information. That is:

$$\frac{1}{2}\,\mathrm{Tr}\left[\delta\rho\,\mathbb{J}_{f,\rho}^{-1}[\delta\rho]\right] = \langle\boldsymbol{\delta},\boldsymbol{\delta}\rangle_{\boldsymbol{\rho}} \quad \text{and} \quad \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{Tr}\left[\delta\rho\,\mathbb{J}_{f,\rho}^{-1}[\delta\rho]\right] = \frac{\mathrm{d}}{\mathrm{d}t}\,\langle\boldsymbol{\delta},\boldsymbol{\delta}\rangle_{\boldsymbol{\rho}}\,. \tag{46}$$

In fact, consider the scenario in which initially the perturbation is of the form $\delta\rho \equiv \delta\rho_\Delta$. In order to compute the derivative, one considers the evolution of the state $\rho+\delta\rho$ for a time $\mathrm{d}t$, which we denote by $\tilde\rho + \delta\tilde\rho$. Then, the perturbation has the form $\delta\tilde\rho = \delta\tilde\rho_\Delta + \mathrm{d}t\,\delta\tilde\rho_C$, so that we are in the situation of Eq. (45) (notice that it doesn't matter whether we take $\delta\tilde\rho_\Delta$ to be diagonal with respect to ρ or to $\tilde\rho$, as this difference only contributes to order $\mathcal{O}(\mathrm{d}t)$). Since quadratic terms in $\mathrm{d}t$ do not contribute to the derivative, these considerations prove Eq. (46).

Finally, in the next Lemma we show that there are special points for which the trace distance and the Fisher information locally coincide. This result allows us to lift the many constructions present in the literature for the trace distance to the study of the Fisher information metric.

Lemma 3. *Choose an arbitrary perturbation $\delta\rho$. Then, consider the state $\rho_{\delta\rho} = \frac{|\delta\rho|}{\mathrm{Tr}[|\delta\rho|]}$. It holds that*

$$D^2_{\mathrm{Fish}}(\rho_{\delta\rho}, \rho_{\delta\rho} + \delta\rho) = \frac{1}{2} D^2_{\mathrm{Tr}}(\rho_{\delta\rho}, \rho_{\delta\rho} + \delta\rho) = \frac{1}{2} \mathrm{Tr}[|\delta\rho|]^2 . \tag{47}$$

Moreover, since $[\rho_{\delta\rho}, \delta\rho] = 0$, one can use Corollary 1 to show that:

$$\frac{\mathrm{d}}{\mathrm{d}t} D^2_{\mathrm{Fish}}(\rho_{\delta\rho}, \rho_{\delta\rho} + \delta\rho) = \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} D^2_{\mathrm{Tr}}(\rho_{\delta\rho}, \rho_{\delta\rho} + \delta\rho) = \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Tr}[|\delta\rho|]^2 \tag{48}$$

In fact, since $[\rho_{\delta\rho}, \delta\rho] = 0$, the quantum Fisher information and its derivative can be studied just by looking at the quantity $\langle \boldsymbol{\delta}, \boldsymbol{\delta} \rangle_{\rho_{\delta\rho}}$. But this is given by:

$$2 \langle \boldsymbol{\delta}, \boldsymbol{\delta} \rangle_{\rho_{\delta\rho}} = \sum_i \frac{\delta_i^2}{\left(\frac{|\delta_i|}{\mathrm{Tr}[|\delta\rho|]} \right)} = \sum_i \frac{\delta_i^2}{|\delta_i|} \sum_j |\delta_j| = \left(\sum_i |\delta_i| \right)^2 = D^2_{\mathrm{Tr}}(\rho_{\delta\rho}, \rho_{\delta\rho} + \delta\rho). \tag{49}$$

The fact that not only one can locally identify the Fisher distance and the trace distance, but also their first derivatives will be of key importance in many of our derivations.

D Additional proofs

D.1 Theorem 1: quantum case

In the case of a quantum map $\mathcal{T}^{(t,0)}$ non-Markovianity means that $\mathcal{T}^{(t,0)}$ is not CP-divisible. This implies that there exists a time t for which the Choi state $\mathcal{T}^{(t+\mathrm{d}t,t)} \otimes \mathbb{1}_N [|\psi^+\rangle\langle\psi^+|]$ develops some negative eigenvalue (where $|\psi^+\rangle$ is the maximally entangled state). Call $|v\rangle$ the corresponding eigenvector. Since $\mathcal{T}^{(t+\mathrm{d}t,t)}$ is infinitesimal, the Choi state must be close to $|\psi^+\rangle\langle\psi^+|$. But then, in order for $|v\rangle$ to correspond to a negative eigenvalue it must contain a non-zero component $|v_\perp\rangle$ orthogonal to $|\psi^+\rangle$. To see this, assume the opposite, i.e., $|v\rangle \equiv |\psi^+\rangle$ (as it is parallel to $|\psi^+\rangle$ and normalised); then, from perturbation theory we know that the corresponding eigenvalue must be $1 + \mathcal{O}(\mathrm{d}t) > 0$. This contradicts the assumption that $|v\rangle$ is associated to a negative eigenvalue.

Consider now the state $\rho = |\psi^+\rangle\langle\psi^+|$ and the perturbation $\delta\rho = \varepsilon(|v_\perp\rangle\langle v_\perp| - |\psi^+\rangle\langle\psi^+|)$. With this choice we have that $[\rho, \delta\rho] = 0$. Moreover, the evolution of $\delta\rho$ through $\mathcal{T}^{(t+\mathrm{d}t)}$ can only generate an off-diagonal component of order $\mathcal{O}(\mathrm{d}t)$. Hence, we are in the situation of Corollary 2, that is we can neglect the off-diagonal contributions completely. In this way, we can simulate the process with a classical dynamics by only considering transitions from the diagonal into itself. Then, it is enough to notice that the classical rate $a_{v_\perp \leftarrow \psi^+}$ is negative by assumption. This concludes the proof.

D.2 Theorem 2: many copy case

Consider the dynamics $\bar{T}^{(t,0)} = T^{(t,0)\otimes n} \otimes \mathbb{1}_M$ acting on the state space $\mathcal{S}(\mathbb{R}^{N^{\otimes n}} \otimes \mathbb{R}^M)$. The rate matrix in this case is given by

$$\bar{R}^{(t)} = \frac{\mathrm{d}}{\mathrm{d}t} \bar{T}^{(t,0)} = \sum_{l=1}^n \mathbb{1}_N^{\otimes(l-1)} \otimes R^{(t)} \otimes \mathbb{1}_N^{\otimes(n-l)} \otimes \mathbb{1}_M . \tag{50}$$

Denoting by latin letters the indexes on the copies of the system space, and by greek letters the indexes of the ancillary space, we can express the rates as

$$\bar{a}^{(t)}_{(i_1 i_2 \dots i_n, \alpha) \leftarrow (j_1 j_2 \dots j_n, \beta)} = \sum_{l=1}^{n} a^{(t)}_{i_l \leftarrow j_l} \left(\delta_{i_1 j_1} \dots \delta_{i_{l-1} j_{l-1}} \delta_{i_{l+1} j_{l+1}} \dots \delta_{i_n j_n} \delta_{\alpha \beta} \right). \tag{51}$$

In analogy with the construction for a single copy, it is now sufficient to consider a dynamics $T^{(t,0)}$ such that the image of $T^{(t,0)}$ is contained in a small ball around an appropriate π, i.e.,

$$\boldsymbol{p}(t) \sim \boldsymbol{\pi}^{\otimes n} \otimes \boldsymbol{q} + \mathcal{O}(\varepsilon), \ \boldsymbol{\pi} \in \mathcal{S}(\mathbb{R}^N), \ \boldsymbol{q} \in \mathcal{S}(\mathbb{R}^M). \tag{52}$$

In this case Eq. (35) takes the form

$$-\sum_{\vec{i} \neq \vec{j}, \alpha} \bar{a}_{\vec{i}\alpha \leftarrow \vec{j}\alpha} \left(\frac{d_{\vec{i}\alpha}}{p_{\vec{i}\alpha}} - \frac{d_{\vec{j}\alpha}}{p_{\vec{j}\alpha}} \right)^2 p_{\vec{j}\alpha} + \mathcal{O}(\varepsilon) \quad \text{with } p_{\vec{j}\alpha} = \pi_{j_1} \dots \pi_{j_n} q_\alpha. \tag{53}$$

Suppose now that $a_{1 \leftarrow 2}$ is the only negative rate of $R^{(t)}$. We want to show that despite the onset of non-Markovianity, the Fisher distance continues to decrease for all the points in the image of $\bar{T}^{(t,0)}$, i.e., the derivative expressed in Eq. (53) is negative. Then, consider for the moment the positive contributions to Eq. (53). These are given by

$$-\sum_{l=1}^{n} a_{1 \leftarrow 2} \left(\frac{d_{\vec{j}_{l=1}\alpha}}{p_{\vec{j}_{l=1}\alpha}} - \frac{d_{\vec{j}_{l=2}\alpha}}{p_{\vec{j}_{l=2}\alpha}} \right)^2 p_{\vec{j}_{l=2}\alpha}, \tag{54}$$

where $\vec{j}_{l=2}$ is the string $\vec{j}_{l=2} = (j_1, \dots, j_{l-1}, 2, j_{l+1}, \dots, j_n)$. We can now compare Eq. (54) with the following contribution

$$-\sum_{l=1}^{n} a_{2 \leftarrow 1} \left(\frac{d_{\vec{j}_{l=1}\alpha}}{p_{\vec{j}_{l=1}\alpha}} - \frac{d_{\vec{j}_{l=2}\alpha}}{p_{\vec{j}_{l=2}\alpha}} \right)^2 p_{\vec{j}_{l=1}\alpha}, \tag{55}$$

where we defined $\vec{j}_{l=1}$ in analogy with $\vec{j}_{l=2}$. Summing up Eq. (54) and Eq. (55) then turns out to be negative whenever $p_{\vec{j}_{l=1}\alpha} a_{2 \leftarrow 1} \geq p_{\vec{j}_{l=2}\alpha} |a_{1 \leftarrow 2}|$, i.e., $\pi_1 a_{2 \leftarrow 1} \geq \pi_2 |a_{1 \leftarrow 2}|$, in complete analogy with what happened in the single copy case.

D.2.1   Theorem 2: quantum case

Theorem 2 holds also in the quantum setting. The extension of the proof makes use again of Lemma 2. In particular, we exploit a quantum dynamical map $\mathcal{T}^{(t,0)}$ with a strong dephasing which reduces the states to be (almost) classical. That is, consider the evolution given

$$\mathcal{T}^{(t,0)}[\rho] = \mathcal{D}^{(\varepsilon_2)} \circ \mathcal{F}^{(\varepsilon_1)}_\pi[\rho], \tag{56}$$

where

$$\mathcal{F}^{(\varepsilon_1)}_\pi[\rho] = (1 - \varepsilon_1)\pi + \varepsilon_1 \rho, \tag{57}$$

$$\mathcal{D}^{(\varepsilon_2)}[\rho] = (1 - \varepsilon_2)\rho_D + \varepsilon_2 \rho, \quad \rho_D = \sum_i |i\rangle\langle i| \rho |i\rangle\langle i|, \tag{58}$$

and $|i\rangle$ is an eigenbasis of π (so that $\mathcal{D}$ is the dephasing operator in the basis of π).

Now thanks to Lemma 2 for small enough ε_2, one can compute the Quantum Fisher Information and its instantaneous variation by reducing it to its classical value. To be precise, as the Theorem considers multiple copies of the channel and ancillary degrees of freedom, one

needs to verify that also in this case the evolution of the Fisher Information collapses onto the classical case. Consider any initially prepared ρ and $\rho + \delta\rho$ quantum states of $\mathcal{H}^{\otimes n} \otimes \mathbb{C}^M$, where $\mathcal{H}$ is the N-dimensional space of the single-copy channel, and an M-dimensional ancilla is allowed. Then the evolution to time t is given by

$$\bar{\mathcal{T}}^{(t,0)} = \mathcal{T}^{(t,0)\otimes n} \otimes \mathbb{1}_M \,, \tag{59}$$

where $\mathcal{T}^{(t,0)}$ is as in Eq. (56). By defining σ to be the reduced state of ρ on the ancillary degrees of freedom, one has

$$\bar{\mathcal{T}}^{(t,0)}[\rho] = \pi^{\otimes n} \otimes \sigma + \mathcal{O}(\varepsilon_1) \,. \tag{60}$$

At the same time, it also holds that

$$\bar{\mathcal{T}}^{(t,0)}[\delta\rho] = \varepsilon_1 \delta\rho_D + \mathcal{O}(\varepsilon_1 \varepsilon_2) \,, \tag{61}$$

where $\delta\rho_D = \mathcal{D}^{(0)\otimes n} \otimes \mathbb{1}_M[\delta\rho]$. As such it can be expressed as

$$\delta\rho_D = \sum_\gamma \theta_D^{(\gamma)} \otimes \omega^{(\gamma)} \,, \quad [\theta_D^{(\gamma)}, \pi^{\otimes n}] = 0 \,, \tag{62}$$

where the $\theta_D^{(\gamma)}$ are operators on $\mathcal{H}^{\otimes n}$ and $\omega^{(\gamma)}$ on $\mathbb{C}^M$. The quantum Fisher Information can then be computed as

$$\frac{\varepsilon_1^2}{2} \mathrm{Tr}\Big[\delta\rho_D \, \mathbb{J}_{f,\pi^{\otimes n}\otimes\sigma}^{-1}[\delta\rho_D] \Big] + \mathcal{O}\big(\varepsilon_1^3\big) + \mathcal{O}\big(\varepsilon_1^2 \varepsilon_2\big) \,. \tag{63}$$

By substituting the expression for $\delta\rho_D$, we get that the leading order is

$$\frac{\varepsilon_1^2}{2} \sum_{\gamma,\gamma'} \mathrm{Tr}\Big[\theta_D^{(\gamma)} \mathbb{J}_{f,\pi^{\otimes n}}^{-1} \theta_D^{(\gamma')} \Big] \mathrm{Tr}\Big[\omega^{(\gamma)} \mathbb{J}_{f,\sigma}^{-1} \omega^{(\gamma')} \Big] := \frac{\varepsilon_1^2}{2} \sum_{\gamma,\gamma'} \mathcal{M}_{\gamma\gamma'}^{(1)} \mathcal{M}_{\gamma\gamma'}^{(2)} \,. \tag{64}$$

Given that only the first trace is time-dependent (the evolution on the ancillary degrees of freedom is trivial), the instantaneous derivative of the above equation can be written as

$$\frac{\varepsilon_1^2}{2} \sum_{\gamma,\gamma'} \Big(\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{M}_{\gamma\gamma'}^{(1)} \Big) \mathcal{M}_{\gamma\gamma'}^{(2)} \,, \tag{65}$$

that is, as the trace-product of two matrices, $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$. We now notice that $\mathcal{M}^{(2)}$ is positive definite, therefore it is enough for $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{M}^{(1)}$ to be negative definite in order for the product to be ≤ 0.

To prove the quantum version of Theorem 2 it is then sufficient to provide a non-Markovian evolution for which $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{M}^{(1)}$ is negative definite. It is also enough to consider the case in which $\mathcal{M}^{(1)}$ reduces to its classical value, due to $[\theta_D^{(\gamma)}, \pi] = 0$. That is

$$\mathcal{M}_{\gamma\gamma'}^{(1)} = 2 \Big\langle \boldsymbol{\theta}^{(\gamma)}, \boldsymbol{\theta}^{(\gamma')} \Big\rangle_\pi \,, \tag{66}$$

where $\boldsymbol{\theta}^{(\gamma)}$ and $\boldsymbol{\pi}$ are the vectors given by the diagonal components of $\theta^{(\gamma)}$ and π respectively. It is then clear that choosing at time t a dynamics of the form

$$\mathcal{T}^{(t+\delta t, t)} = \mathbb{1} + \delta t \mathcal{L} \,, \tag{67}$$

where $\mathcal{L}$ is a semi-classical Lindbladian with rates $a_{i \leftarrow j}$ and jump operators $|i\rangle\langle j|$

$$\mathcal{L}[\rho] = \sum_{i \neq j} a_{i \leftarrow j} \left(|i\rangle\langle j| \rho |j\rangle\langle i| - \frac{1}{2}\{|j\rangle\langle j|, \rho\} \right), \tag{68}$$

will induce the classical dynamics locally described by the rates $a_{i \leftarrow j}$ on the diagonal subspace of quantum states.

Finally, the classical version of Theorem 2, given above D.2, ensures that there are instances in which at least one of the $a_{i \leftarrow j}$ is negative (and hence $\mathcal{T}^{t+\delta t,t}$ non-CP) while $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{M}^{(1)}$ is negative definite. This concludes the proof.

D.3 Proof of Theorem 3

Despite the negative result given by Theorem 2, we provide here an explicit construction to convert a non-Markovianity witness based on the trace-distance to one that uses the Fisher metric by exploiting a particular kind of post-processing.

Consider, in fact, the case in which during a non-Markovian dynamics there is some $\delta\rho$ such that $\frac{\mathrm{d}}{\mathrm{d}t}|\delta\rho|^2_{\mathrm{Tr}}$ is increasing in time. As we saw in Appendix A, to guarantee the existence of such $\delta\rho$, it is sufficient to use an ancilla, namely by considering the dynamics $T^{(t,0)} \otimes \mathbb{1}_2$ for the classical case, and $\mathcal{T}^{(t,0)} \otimes \mathbb{1}_{N+1}$ for the quantum case.

Thanks to Lemma 3, we also know that on the base-point $\rho_{\delta\rho}$ it holds that

$$\frac{\mathrm{d}}{\mathrm{d}t} D_{\mathrm{Fish}}(\rho_{\delta\rho}, \rho_{\delta\rho} + \delta\rho) = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|\delta\rho|^2_{\mathrm{Tr}} > 0. \tag{69}$$

It is useful to repeat the computation of this derivative here. We obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} D_{\mathrm{Fish}}(\rho_{\delta\rho}, \rho_{\delta\rho} + \delta\rho) = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} \sum_i \frac{\delta\rho_i^2}{(\rho_{\delta\rho})_i} = \sum_i \left(\frac{\delta\rho_i \delta\dot\rho_i}{(\rho_{\delta\rho})_i} - \frac{1}{2}\frac{\delta\rho_i^2 (\dot\rho_{\delta\rho})_i}{(\rho_{\delta\rho})_i^2} \right). \tag{70}$$

The last term does not contribute to the sum: in fact, the first term already gives $\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|\delta\rho|^2_{\mathrm{Tr}}$, so the last term must be zero due to Lemma 3. On the other hand, it is not difficult to carry out the explicit calculation, giving

$$\sum_i \frac{\delta\rho_i^2 (\dot\rho_{\delta\rho})_i}{(\rho_{\delta\rho})_i^2} = \left(\sum_i \frac{\delta\rho_i^2 (\dot\rho_{\delta\rho})_i}{\delta\rho_i^2} \right) \left(\sum_j |\delta\rho_j| \right)^2 = 0, \tag{71}$$

where we used the expression of $\rho_{\delta\rho}$, together with the fact that $\dot\rho_{\delta\rho}$ is traceless. Hence, we showed that despite the explicit dependence of the Fisher information on $\rho_{\delta\rho}$, there is no contribution in its derivative coming from $\dot\rho_{\delta\rho}$. In some sense, we can deduce that the base-point could be *frozen* and the Fisher information would still have a backflow.

This intuition inspires the following construction. Define the stochastic operator F as

$$F_\pi^{(\varepsilon)}[\sigma] = (1 - \varepsilon)\pi + \varepsilon\sigma, \tag{72}$$

where π is some arbitrary state and ε is a small parameter. That is, the operator F is an almost-complete erasure of prior information, sending all the states close to π. If σ and τ are two states, their difference $\Delta := \sigma - \tau$ is transformed as:

$$F_\pi^{(\varepsilon)}[\Delta] = F_\pi^{(\varepsilon)}[\sigma] - F_\pi^{(\varepsilon)}[\tau] = \varepsilon\Delta. \tag{73}$$

In particular, the Fisher information between $\boldsymbol{p}$ and $\boldsymbol{p} + \boldsymbol{d}$ transforms under the application of this filter as

$$D^2_{\text{Fish}}(F^{(\varepsilon)}_\pi[\boldsymbol{p}], F^{(\varepsilon)}_\pi[\boldsymbol{p} + \boldsymbol{d}]) = \sum_i \varepsilon^2 \frac{d_i^2}{\pi_i} + \mathcal{O}\left(\varepsilon^3\right) \simeq D^2_{\text{Fish}}(\pi, \pi + \varepsilon \boldsymbol{d}) + \mathcal{O}\left(\varepsilon^3\right). \tag{74}$$

Then, choose π as $\pi := \rho_{\delta\rho}$. Then, we can construct a witness of the non-Markovianity of the evolution at time t through the Fisher information, by applying the filter $F^{(\varepsilon)}_{\rho_{\delta\rho}}$ to the final state. This gives:

$$\frac{\mathrm{d}}{\mathrm{d}t} D^2_{\text{Fish}}(F^{(\varepsilon)}_{\rho_{\delta\rho}}[\rho(t)], F^{(\varepsilon)}_{\rho_{\delta\rho}}[\rho(t) + \delta\rho(t)]) = \tag{75}$$

$$= \lim_{\mathrm{d}t \to 0} \left(\frac{D^2_{\text{Fish}}(F^{(\varepsilon)}_{\rho_{\delta\rho}}[\rho(t+\mathrm{d}t)], F^{(\varepsilon)}_{\rho_{\delta\rho}}[\rho(t+\mathrm{d}t) + \delta\rho(t+\mathrm{d}t)]) - D^2_{\text{Fish}}(F^{(\varepsilon)}_{\rho_{\delta\rho}}[\rho(t)], F^{(\varepsilon)}_{\rho_{\delta\rho}}[\rho(t) + \delta\rho(t)])}{\mathrm{d}t} \right) = $$

$$\tag{76}$$

$$= \lim_{\mathrm{d}t \to 0} \left(\frac{D^2_{\text{Fish}}(\rho_{\delta\rho}, \rho_{\delta\rho} + \varepsilon \, \delta\rho(t+\mathrm{d}t)) - D^2_{\text{Fish}}(\rho_{\delta\rho}, \rho_{\delta\rho} + \varepsilon \, \delta\rho(t)) + \mathcal{O}\left(\varepsilon^3\right)}{\mathrm{d}t} \right) = \tag{77}$$

$$= \varepsilon^2 \frac{\mathrm{d}}{\mathrm{d}t} |\delta\rho(t)|^2_{\text{Tr}} + \mathcal{O}\left(\varepsilon^3\right), \tag{78}$$

where in Eq. (77) we used the result from Eq. (74). Regarding the last equality, on the other hand, it should be noticed that in Eq. (77) $\rho_{\delta\rho}$ does not depend on time. Still, thanks to the remarks made above, we know that $\dot{\rho}_{\delta\rho}$ does not contribute to the derivative, so we can apply Lemma 3. Incidentally, in the main text we use the notation $F_{\boldsymbol{d}} := F^{(\varepsilon)}_{\rho_{\delta\rho}}$, where $\boldsymbol{d}$ is the diagonal part of $\delta\rho$.

Now, the quantity constructed is contractive under Markovian dynamics, as the trace distance is contractive. On the other hand, by choosing $\delta\rho$ to be a witness of non-Markovianity for the trace distance, we also obtain a witness in this scenario. There is a shortcoming to this construction, though: in the definition of the post-processing $F^{(\varepsilon)}_{\rho_{\delta\rho}}$ one uses the perturbation $\delta\rho(t)$, and not just $\delta\rho(0)$ (in fact, to be more precise, the filter is actually given by $F^{(\varepsilon)}_{\rho_{\delta\rho(t)}}$). In this way, one has to know in advance the structure of the dynamics to provide an explicit construction. Still, this example serves more as a proof of principle of the possibility of designing post-processing filters to exploit the Fisher information for the detection of non-Markovianity.

References

[1] H.-P. Breuer, F. Petruccione *et al.*, *The theory of open quantum systems*, Oxford University Press on Demand, doi:10.1093/acprof:oso/9780199213900.001.0001 (2002).

[2] A. Rivas and S. F. Huelga, *Open quantum systems*, vol. 10, Springer, doi:https://doi.org/10.1007/978-3-642-23354-8 (2012).

[3] Á. Rivas, S. F. Huelga and M. B. Plenio, *Quantum non-markovianity: characterization, quantification and detection*, Reports on Progress in Physics **77**(9), 094001 (2014), doi:10.1088/0034-4885/77/9/094001.

[4] H.-P. Breuer, E.-M. Laine, J. Piilo and B. Vacchini, *Colloquium: Non-markovian dynamics in open quantum systems*, Rev. Mod. Phys. **88**, 021002 (2016), doi:10.1103/RevModPhys.88.021002.

[5] B. Bylicka, M. Johansson and A. Acín, *Constructive method for detecting the information backflow of non-markovian dynamics*, Phys. Rev. Lett. **118**, 120501 (2017), doi:10.1103/PhysRevLett.118.120501.

[6] F. Buscemi and N. Datta, *Equivalence between divisibility and monotonic decrease of information in classical and quantum stochastic processes*, Phys. Rev. A **93**, 012101 (2016), doi:10.1103/PhysRevA.93.012101.

[7] B. Bylicka, D. Chruściński and S. Maniscalco, *Non-markovianity and reservoir memory of quantum channels: a quantum information theory perspective*, Scientific Reports **4** (2014), doi:https://doi.org/10.1038/srep05720.

[8] S. Lorenzo, F. Plastina and M. Paternostro, *Geometrical characterization of non-markovianity*, Phys. Rev. A **88**, 020102 (2013), doi:10.1103/PhysRevA.88.020102.

[9] D. D. Santis and M. Johansson, *Equivalence between non-markovian dynamics and correlation backflows*, New Journal of Physics **22**(9), 093034 (2020), doi:10.1088/1367-2630/abaf6a.

[10] J. Kołodyński, S. Rana and A. Streltsov, *Entanglement negativity as a universal non-markovianity witness*, Phys. Rev. A **101**, 020303 (2020), doi:10.1103/PhysRevA.101.020303.

[11] M. A. Nielsen and I. Chuang, *Quantum computation and quantum information*, doi:doi:10.1017/CBO9780511976667 (2002).

[12] W. K. Wootters, *Statistical distance and hilbert space*, Phys. Rev. D **23**, 357 (1981), doi:10.1103/PhysRevD.23.357.

[13] H. Cramer, *Mathematical methods of statistics*, Princeton University Press, ISBN 0691005478, doi:https://doi.org/10.1515/9781400883868 (1946).

[14] C. R. Rao, *Information and the accuracy attainable in the estimation of statistical parameters*, Kotz S., Johnson N.L. (eds) Breakthroughs in Statistics. Springer Series in Statistics (Perspectives in Statistics) (1992), doi:10.1007/978-1-4612-0919-5.

[15] H. Chernoff, *A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the sum of Observations*, The Annals of Mathematical Statistics **23**(4), 493 (1952), doi:10.1214/aoms/1177729330.

[16] S. Kullback, *Information theory and statistics*, Dover. Publications, ISBN 978-0486696843 (1997).

[17] I. Csiszár and P. C. Shields, *Information Theory and Statistics: A Tutorial*, Foundations and Trends® in Communications and Information Theory **1**(4), 417 (2004), doi:10.1561/0100000004.

[18] X.-M. Lu, X. Wang and C. Sun, *Quantum fisher information flow and non-markovian processes of open systems*, Physical Review A **82**(4), 042103 (2010), doi:https://doi.org/10.1103/PhysRevA.82.042103.

[19] N. N. Cencov, *Statistical decision rules and optimal inference*, 53. American Mathematical Soc., doi:https://doi.org/10.1090/mmono/053 (2000).

[20] L. L. Campbell, *An extended čencov characterization of the information metric*, Proceedings of the American Mathematical Society **98**(1), 135 (1986), doi:https://doi.org/10.2307/2045782.

[21] D. Petz, *Monotone metrics on matrix spaces*, Linear Algebra and its Applications **244**, 81 (1996), doi:10.1016/0024-3795(94)00211-8.

[22] M. J. Hall, J. D. Cresser, L. Li and E. Andersson, *Canonical form of master equations and characterization of non-markovianity*, Physical Review A **89**(4), 042120 (2014), doi:10.1103/PhysRevA.89.042120.

[23] F. Buscemi and V. Scarani, *Fluctuation theorems from bayesian retrodiction*, Phys. Rev. E **103**, 052111 (2021), doi:10.1103/PhysRevE.103.052111.

[24] J. Surace and M. Scandi, *State retrieval beyond bayes' retrodiction and reverse processes*, doi:10.48550/ARXIV.2201.09899 (2022), arXiv:2201.09899.

[25] M. Scandi, P. Abiuso, J. Surace and D. De Santis, Quantum Fisher Information: divisible dynamics, reverse processes and witnesses for Non-Markovianity *(In preparation)* .

[26] A. Lesniewski and M. B. Ruskai, *Monotone Riemannian Metrics and Relative Entropy on Non-Commutative Probability Spaces*, Journal of Mathematical Physics **40**(11), 5702 (1999), doi:10.1063/1.533053, math-ph/9808016.