2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Backward Map for Filter Stability Analysis

Jin Won Kim, Anant A. Joshi and Prashant G. Mehta

2024 IEEE 63rd Conference on Decision and Control (CDC) | 979-8-3503-1633-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/CDC56724.2024.10886772

***Abstract*— A backward map is introduced for the purposes of analysis of nonlinear (stochastic) filter stability. The backward map is important because filter-stability, in the sense of χ^2-divergence, follows from a certain variance decay property associated with the backward map. To show this property requires additional assumptions on the hidden Markov model (HMM). The analysis in this paper is based on introducing a Poincaré Inequality (PI) for HMMs with white noise observations. In finite state-space settings, PI is related to both the ergodicity of the Markov process as well as the observability of the HMM. It is shown that the Poincaré constant is positive if and only if the HMM is detectable.**

I. Introduction

Dissipation is at the heart of any stability theory for dynamical systems. For Markov processes, dissipation is referred to as *variance decay*. To illustrate the key ideas, consider a Feller-Markov process $X = \{X_t : t \geq 0\}$ taking values in a Polish state-space $\mathbb{S}$ and suppose $\bar{\mu}$ is a given invariant measure. The fundamental object of interest is the Markov semigroup defined by [1, Eq. (1.1.1)]

$$(P_t f)(x) := \mathsf{E}^x(f(X_t)), \quad x \in \mathbb{S}, \quad t \geq 0$$

for $f : \mathbb{S} \to \mathbb{R}$ in some class of measurable functions. The problem of stochastic stability is to show that $P_t f \to \bar{\mu}(f)$, in some suitable sense, as $t \to \infty$. As defined, $(P_t f)(x)$ has an interpretation as the expectation of the random variable $f(X_t)$ starting from an initial condition $X_0 = x$. Therefore, $P_t f \to \bar{\mu}(f)$ means that this expectation asymptotically converges to its stationary value $\bar{\mu}(f)$ for all choices (in suitable sense) of the initial conditions $x \in \mathbb{S}$.

The dissipation equation requires the notation,

$$\begin{aligned} &\text{(variance)} \qquad \mathcal{V}^{\bar{\mu}}(f) := \bar{\mu}(f^2) - \bar{\mu}(f)^2 \\ &\text{(energy)} \qquad \mathcal{E}^{\bar{\mu}}(f) := \bar{\mu}(\Gamma f) \end{aligned}$$

where Γ is the so called carré du champ operator (Defn. 2). The operator is a positive-definite bilinear form. Using these definitions, the dissipation equation arises as

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{V}^{\bar{\mu}}(P_t f) = -\mathcal{E}^{\bar{\mu}}(P_t f), \quad t \geq 0$$

The calculation for the same appears in Appendix A, see also [1, Thm. 4.2.5.)]. The equation shows that $\{\mathcal{V}^{\bar{\mu}}(P_t f) : t \geq 0\}$ is non-increasing. To show that the variance decays to zero requires a suitable relationship between energy and variance. The simplest such relationship is through the Poincaré Inequality (PI):

$$\text{(PI)} \qquad \mathcal{E}^{\bar{\mu}}(f) \geq c\,\mathcal{V}^{\bar{\mu}}(f), \quad \forall f \in \mathcal{D}$$

where the sharpest such constant c is referred to as the Poincaré constant; here, $\mathcal{D}$ is a suitable space of test functions for which the energy $\mathcal{E}^{\bar{\mu}}(f)$ is well-defined (see Defn. 2). PI is useful to conclude stochastic stability where c gives the exponential rate of convergence. For reversible Markov processes, c is referred to as the spectral gap constant.

A. Aims and contributions of this paper

This paper is concerned with extension of variance decay to the study of the nonlinear filter. Specifically, the following questions are of interest:

- **Q1.** What is the appropriate notion of variance decay for a nonlinear filter? And how is it related to filter stability?
- **Q2.** What is the appropriate generalization of the dissipation equation for the nonlinear filter?
- **Q3.** What is the appropriate generalization of the Poincaré inequality for the filter? And how is it related to the hidden Markov model (HMM) properties such as observability and detectability?

In this paper, we provide an answer to each of these questions (see Prop. 1 for Q1, Prop. 3 for Q2, and Prop. 6-7 for Q3). An original contribution of this paper is the backward map that is introduced here for a general class of HMMs. The backward map is important because filter-stability, in the sense of χ^2-divergence, follows from a certain variance decay property associated with the backward map. While the backward map and the variance decay is for a general class of HMMs, the answers to Q2 and Q3 are given for HMM with white noise observations. The overall approach may be regarded as an *optimal control approach* to filter stability based on our recent work on duality [2], [3], [4]. Our approach is contrasted with the intrinsic approach to filter stability based upon specification of a certain forward map [5].

B. Outline of the remainder of this paper

Sec. II contains math preliminaries for the HMM and the filter stability problem. The backward map is introduced in Sec. III and specialized to white noise observations in Sec. IV. For this HMM, the definition of the Poincaré Inequality (PI) is introduced in Sec. V. The PI is related to the HMM model properties in Sec. VI (for the finite state-space settings) and illustrated using numerics in Sec. VII. The proofs appear in the Appendix.

This work is supported in part by the AFOSR award FA9550-23-1-0060, the NSF award 2336137 (Joshi, Mehta) and the DFG grant 318763901/SFB1294 (Kim).

J. W. Kim is with the Department of Mechanical and System Design Engineering at the Hongik University. (e-mail: jin.won.kim@hongik.ac.kr)

A. A. Joshi and P. G. Mehta are with the Coordinated Science Laboratory and the Department of Mechanical Science and Engineering at the University of Illinois at Urbana-Champaign. (e-mail: anantaj2; mehtapg@illinois.edu)

II. Mathematical Preliminaries

A. Hidden Markov model (HMM)

For the definition and analysis of the nonlinear filter, a standard model for HMM is specified (see [6, Sec. 2]) through construction of a pair of stochastic processes $(X,Z) := \{(X_t, Z_t) : 0 \le t \le T\}$ on probability space $(\Omega, \mathcal{F}_T, \mathsf{P})$ as follows:

- The state-space $\mathbb{S}$ is a locally compact Polish space. The important examples are (i) $\mathbb{S} = \{1, 2, \cdots, d\}$ of finite or countable cardinality, and (ii) $\mathbb{S} \subseteq \mathbb{R}^d$.
- The observation-space $\mathbb{O} = \mathbb{R}^m$.
- The signal-observation process (X,Z) is a Feller-Markov process.
- The state process X is a Feller-Markov process with $X_0 \sim \mu \in \mathcal{P}(\mathbb{S})$. Here, $\mathcal{P}(\mathbb{S})$ is the space of probability measures defined on the Borel σ-algebra on $\mathbb{S}$ and μ is referred to as the prior.
- The observation process Z has $Z_0 = 0$ and conditionally independent increments given the state process X. That is, given X_t, an increment $Z_s - Z_t$ is independent of $\mathcal{Z}_t := \sigma\big(\{Z_s : 0 \le s \le t\}\big)$, for all $s > t$. The filtration generated by the observations is denoted $\mathcal{Z} := \{\mathcal{Z}_t : 0 \le t \le T\}$.

The objective of nonlinear filtering is to compute the conditional expectation

$$\pi_T(f) := \mathsf{E}\big(f(X_T) \mid \mathcal{Z}_T\big), \quad f \in C_b(\mathbb{S})$$

where $C_b(\mathbb{S})$ is the space of continuous and bounded functions. The conditional measure π_T is referred to as the *nonlinear filter*.

To stress the dependence on the prior μ, a standard convention is to denote the probability space as $(\Omega, \mathcal{F}_T, \mathsf{P}^\mu)$, the expectation operator as E^μ, and the nonlinear filter as π_T^μ. In practice, the prior may not be known. With an incorrect choice of prior $\nu \in \mathcal{P}(\mathbb{S})$, the filter is denoted as π_T^ν. The precise meaning for these along with the definition of filter stability appear next.

B. Filter stability

Let $\rho \in \mathcal{P}(\mathbb{S})$. On the common measurable space $(\Omega, \mathcal{F}_T)$, P^ρ is used to denote another probability measure such that the transition law of (X,Z) is identical but $X_0 \sim \rho$ (see [7, Sec. 2.2] for an explicit construction of P^ρ as a probability measure over paths on $\mathbb{S} \times \mathbb{O}$.). The associated expectation operator is denoted by $\mathsf{E}^\rho(\cdot)$ and the nonlinear filter by $\pi_t^\rho(f) = \mathsf{E}^\rho\big(f(X_t) \mid \mathcal{Z}_t\big)$. The two important choices for ρ are as follows:

- $\rho = \mu$. The measure μ has the meaning of the true prior.
- $\rho = \nu$. The measure ν has the meaning of an incorrect prior that is used to compute the filter.

The relationship between P^μ and P^ν is as follows ($\mathsf{P}^\mu|_{\mathcal{Z}_t}$ denotes the restriction of P^μ to the σ-algebra $\mathcal{Z}_t$):

Lemma 1 (Lemma 2.1 in [7]): Suppose $\mu \ll \nu$. Then

- $\mathsf{P}^\mu \ll \mathsf{P}^\nu$, and the change of measure is given by

$$\frac{\mathrm{d}\mathsf{P}^\mu}{\mathrm{d}\mathsf{P}^\nu}(\omega) = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\big(X_0(\omega)\big) \quad \mathsf{P}^\nu\text{-a.s. } \omega$$

- For each $t > 0$, $\pi_t^\mu \ll \pi_t^\nu$, $\mathsf{P}^\mu|_{\mathcal{Z}_t}$-a.s..

Suppose $\mu \ll \nu$. Then $\pi_T^\mu \ll \pi_T^\nu$ from Lem. 1. Denote the Radon-Nikodyn (R-N) derivative as

$$\gamma_T(x) := \frac{\mathrm{d}\pi_T^\mu}{\mathrm{d}\pi_T^\nu}(x), \quad x \in \mathbb{S}$$

It is noted that while $\gamma_0 = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}$ is a deterministic function on $\mathbb{S}$, γ_T is a $\mathcal{Z}_T$-measurable function on $\mathbb{S}$. A filter is said to be stable if the random function $\gamma_T \to 1$, in a suitable sense, as $T \to \infty$. In this paper, the following notion of filter stability is adopted based on χ^2-divergence[1]:

Definition 1: The nonlinear filter is *stable* in the sense of

$$(\chi^2 \text{ divergence}) \qquad \mathsf{E}^\mu\big(\chi^2(\pi_T^\mu \mid \pi_T^\nu)\big) \longrightarrow 0$$

as $T \to \infty$ for every $\mu, \nu \in \mathcal{P}(\mathbb{S})$ such that $\mu \ll \nu$.

Remark 1: $\gamma_T : \mathbb{S} \to \mathbb{R}$ is a non-negative random function on $\mathbb{S}$ with $\mathsf{E}^\nu(\gamma_T(X_T)|\mathcal{Z}_T) = \pi_T^\nu(\gamma_T) = \int_{\mathbb{S}} \gamma_T(x)\,\mathrm{d}\pi_T^\nu(x) = 1$. The square of the function γ_T is denoted by γ_T^2 (That is, $\gamma_T^2(x) := (\gamma_T(x))^2$ for $x \in \mathbb{S}$). Then

$$\mathsf{E}^\nu(|\gamma_T(X_T) - 1|^2|\mathcal{Z}_T) = \pi_T^\nu(\gamma_T^2) - 1 = \chi^2\big(\pi_T^\mu \mid \pi_T^\nu\big)$$

Therefore, χ^2-divergence $\chi^2\big(\pi_T^\mu \mid \pi_T^\nu\big)$ has the meaning of the conditional variance of the random variable $\gamma_T(X_T)$. Next, $\mathsf{E}^\nu(\gamma_T(X_T)) = \mathsf{E}^\nu(\mathsf{E}^\nu(\gamma_T(X_T)|\mathcal{Z}_T)) = 1$ and therefore the variance of $\gamma_T(X_T)$ is given by,

$$\mathrm{var}^\nu(\gamma_T(X_T)) = \mathsf{E}^\nu(|\gamma_T(X_T) - 1|^2)$$

III. Backward map for the nonlinear filter

A key original concept introduced in this paper is the *backward map* $\gamma_T \mapsto y_0$ defined as follows:

$$y_0(x) := \mathsf{E}^\nu(\gamma_T(X_T)|[X_0 = x]), \quad x \in \mathbb{S} \tag{1}$$

The function $y_0 : \mathbb{S} \to \mathbb{R}$ is deterministic, non-negative, and $\nu(y_0) = \mathsf{E}^\nu(\gamma_T(X_T)) = 1$.

Since $\mu \ll \nu$, it follows $\mu(y_0) = \mathsf{E}^\mu\big(\gamma_T(X_T)\big)$. Using the tower property,

$$\mu(y_0) = \mathsf{E}^\mu\big(\gamma_T(X_T)\big) = \mathsf{E}^\mu\big(\mathsf{E}^\mu(\gamma_T(X_T)|\mathcal{Z}_T)\big) = \mathsf{E}^\mu\big(\pi_T^\mu(\gamma_T)\big)$$

Now, $\pi_T^\mu(\gamma_T) = \pi_T^\nu(\gamma_T^2)$, and therefore,

$$\mu(y_0) = \mathsf{E}^\mu\big(\pi_T^\nu(\gamma_T^2)\big)$$

Noting $\pi_T^\nu(\gamma_T^2) - 1 = \chi^2\big(\pi_T^\mu \mid \pi_T^\nu\big)$ is the χ^2-divergence (see formula in Rem. 1),

$$\mathsf{E}^\mu\big(\chi^2\big(\pi_T^\mu \mid \pi_T^\nu\big)\big) = \mu(y_0) - \nu(y_0)$$

Therefore, filter stability in the sense of χ^2-divergence is equivalent to showing that $\mu(y_0) \overset{(T\to\infty)}{\longrightarrow} 1$.

Because $\mu(y_0) - \nu(y_0) = \nu\big((y_0 - 1)(\gamma_0 - 1)\big)$, upon using the Cauchy-Schwarz inequality,

$$|\mathsf{E}^\mu\big(\chi^2(\pi_T^\mu \mid \pi_T^\nu)\big)|^2 \le \mathrm{var}^\nu(y_0(X_0))\,\chi^2(\mu|\nu) \tag{2}$$

where $\mathrm{var}^\nu(y_0(X_0)) = \mathsf{E}^\nu\big(|y_0(X_0) - 1|^2\big)$.

[1]For any two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{S})$ such that $\mu \ll \nu$, the χ^2-divergence $\chi^2(\mu \mid \nu) := \int_{\mathbb{S}}(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x) - 1)^2\,\mathrm{d}\nu(x)$.

From (2), for all such choices of priors $\mu \in \mathcal{P}(\mathbb{S})$ such that $\chi^2(\mu|\nu) < \infty$, a sufficient condition for filter stability is the following:

$$\textbf{(variance decay prop.)} \quad \text{var}^\nu\big(y_0(X_0)\big) \overset{(T\to\infty)}{\longrightarrow} 0 \tag{3}$$

We have thus shown the following:

*Proposition 1 (**Answer to Q1 in Sec. I**):* Consider the backward map $\gamma_T \mapsto y_0$ defined by (1). Suppose $\chi^2(\mu|\nu) < \infty$ and the variance decay property (3) holds. Then the filter is stable in the sense of χ^2-divergence.

Next, from (1), $(y_0(X_0) - 1) = \mathsf{E}^\nu\big((\gamma_T(X_T) - 1)|X_0\big)$, and using Jensen's inequality,

$$\text{var}^\nu\big(y_0(X_0)\big) \leq \text{var}^\nu\big(\gamma_T(X_T)\big) \tag{4}$$

where $\text{var}^\nu(\gamma_T(X_T)) := \mathsf{E}^\nu\big(|\gamma_T(X_T) - 1|^2\big)$. Therefore, the backward map $\gamma_T \mapsto y_0$ is non-expansive: The variance of the random variable $y_0(X_0)$ is no larger than the variance of the random variable $\gamma_T(X_T)$.

Because contractive operators are a subset of non-expansive operators, one may ask if filter stability is obtained from showing that the backward map $\gamma_T \mapsto y_0$ is contractive? The answer is provided in the following proposition:

Proposition 2: Suppose a stronger form of (4) holds s.t.

$$\text{(Assumption)} \quad \text{var}^\nu\big(y_0(X_0)\big) \leq e^{-cT}\text{var}^\nu\big(\gamma_T(X_T)\big) \tag{5}$$

(Because of (4), this is always true with $c = 0$). Then

$$\mathsf{E}^\mu\big(\chi^2(\pi_T^\mu \mid \pi_T^\nu)\big) \leq \frac{1}{\underline{a}} e^{-cT}\chi^2(\mu|\nu)$$

where $\underline{a} = \text{essinf}_{x\in\mathbb{S}}\gamma_0(x)$.

Proof: See Appendix B. ■

Based on the backward map, the analysis of filter stability involves the following objectives:

1) To justify the stronger form (5) under a suitable definition of the Poincaré inequality (PI).
2) Relate PI to the model properties, namely, (i) ergodicity of the Markov process; and (ii) observability/detectability of the HMM.

While the general case remains open, these objectives are described for the special class of HMMs with white noise observations.

The following remark is included to help relate the approach of this paper to the literature. The reader may choose to skip ahead to Sec. IV without any loss of continuity.

Remark 2 (Comparison with the forward map): The backward map is contrasted with the *forward map*, which is the starting point of the intrinsic approach to the problem of filter stability [5]. The *forward map* $\gamma_0 \mapsto \gamma_T$ is defined as follows:

$$\gamma_T(x) = \mathsf{E}^\nu\Big(\frac{\gamma_0(X_0)}{\mathsf{E}^\nu(\gamma_0(X_0) \mid \mathcal{Z}_T)}\Big| \mathcal{Z}_T \vee [X_T = x]\Big), \quad x \in \mathbb{S}$$

Upon using the map to express the total variation, one can show (see [4, Sec. 6.5] for a complete derivation),

$$\lim_{T\to\infty} \mathsf{E}^\mu\big(\|\pi_T^\mu - \pi_T^\nu\|_{\text{TV}}\big) = \mathsf{E}^\nu\Big(\Big|\mathsf{E}^\nu\big(\gamma_0(X_0) \mid \bigcap_{T\geq 0} \mathcal{Z}_\infty \vee \mathcal{F}^X_{[T,\infty)}\big) - \mathsf{E}^\nu\big(\gamma_0(X_0) \mid \mathcal{Z}_\infty\big)\Big|\Big)$$

where $\mathcal{Z}_\infty = \bigcup_{T>0}\mathcal{Z}_T$, $\mathcal{F}^X_{[T,\infty)} = \sigma\big(\{X_t : t \geq T\}\big)$ is the tail sigma-algebra of the state process X. As a function of T, $\mathcal{Z}_\infty \vee \mathcal{F}^X_{[T,\infty)}$ is a decreasing filtration and $\mathcal{Z}_T$ is an increasing filtration. Therefore, by the martingale convergence theorem, both terms on the right-hand side converge as $T \to \infty$. The limit is zero if the following tail sigma-field identity holds:

$$\bigcap_{T\geq 0} \mathcal{Z}_\infty \vee \mathcal{F}^X_{[T,\infty)} \overset{?}{=} \mathcal{Z}_\infty$$

This identity is referred to as the central problem in the stability analysis of the nonlinear filter [8]. The problem generated significant attention (see [9] and references therein).

IV. Embedding the Backward Map in a BSDE

A. White noise observation model

In the remainder of this paper, the observation process Z is according to the stochastic differential equation (SDE):

$$Z_t = \int_0^t h(X_s)\,\mathrm{d}s + W_t, \quad t \geq 0 \tag{6}$$

where $h : \mathbb{S} \to \mathbb{R}^m$ is referred to as the observation function and $W = \{W_t : 0 \leq t \leq T\}$ is an m-dimensional Brownian motion (B.M.). We write W is P-B.M. It is assumed that W is independent of X.

For the ensuing analysis, we also need to specify additional notation for the Markov process X. Specifically, the infinitesimal generator of the Markov process X is denoted by $\mathcal{A}$. In terms of $\mathcal{A}$, an important definition is as follows:

Definition 2 (Defn. 1.4.1. in [1]): The bilinear operator

$$\Gamma(f,g)(x) := (\mathcal{A}fg)(x) - f(x)(\mathcal{A}g)(x) - g(x)(\mathcal{A}f)(x), \; x \in \mathbb{S}$$

defined for every $(f,g) \in \mathcal{D}\times\mathcal{D}$ is called the *carré du champ operator* of the Markov generator $\mathcal{A}$. Here, $\mathcal{D}$ is a vector space of (test) functions that are dense in a suitable L^2 space, stable under products (i.e., $\mathcal{D}$ is an algebra), and $\Gamma : \mathcal{D}\times\mathcal{D} \to \mathcal{D}$ such that $\Gamma(f,f) \geq 0$ for every $f \in \mathcal{D}$. For the case where an invariant measure $\bar{\mu} \in \mathcal{P}(\mathbb{S})$ is available then the natural L^2 space is with respect to the invariant measure: $L^2(\bar{\mu}) = \{f : \mathbb{S} \to \mathbb{R} : \bar{\mu}(f^2) < \infty\}$. We use the notation $(\Gamma f) := \Gamma(f,f)$.

The above is referred to as the *white noise observation model* of nonlinear filtering. The model is denoted by $(\mathcal{A}, h)$.

Because these were stated piecemeal, the main assumptions are stated as follows:

Assumption 1: Consider HMM $(\mathcal{A}, h)$.

1) X is a Feller-Markov process with generator $\mathcal{A}$ and carré du champ Γ.
2) Z is according to the SDE (6) such that the Novikov's condition holds: $\mathsf{E}\Big(\exp\big(\frac{1}{2}\int_0^T |h(X_t)|^2\mathrm{d}t\big)\Big) < \infty$.
3) $\mu, \nu \in \mathcal{P}(\mathbb{S})$ are two priors with $\mu \ll \nu$.

For the HMM $(\mathcal{A},h)$, the nonlinear filter solves the celebrated *Kushner-Stratonovich equation*:

$$\mathrm{d}\pi_t(f) = \pi_t(\mathcal{A}f)\,\mathrm{d}t + \big(\pi_t(hf) - \pi_t(f)\pi_t(h)\big)^{\mathrm{T}}\,\mathrm{d}I_t \qquad (7)$$

where the *innovation process* is defined by

$$I_t := Z_t - \int_0^t \pi_s(h)\,\mathrm{d}s, \quad t \ge 0$$

With $\pi_0 = \rho \in \mathcal{P}(\mathbb{S})$, the filter $\{\pi_t^\rho : 0 \le t \le T\}$ is the solution of (7). Therefore, for the HMM $(\mathcal{A},h)$, γ_T is the R-N ratio of the solution of (7), π_T^μ and π_T^ν, with the two choices of priors, $\pi_0 = \mu$ and $\pi_0 = \nu$, respectively.

B. Embedding the backward map in a BSDE

We continue the analysis of the backward map $\gamma_T \mapsto y_0$ introduced as (1) in Sec. III. For this purpose, consider the backward stochastic differential equation (BSDE):

$$\begin{aligned} -\mathrm{d}Y_t(x) &= \big((\mathcal{A}Y_t)(x) + h^{\mathrm{T}}(x)V_t(x)\big)\,\mathrm{d}t - V_t^{\mathrm{T}}(x)\,\mathrm{d}Z_t, \\ Y_T(x) &= \gamma_T(x), \ x \in \mathbb{S}, \ 0 \le t \le T \end{aligned} \qquad (8)$$

Here $(Y,V) = \{(Y_t(x),V_t(x)) : \Omega \to \mathbb{R}\times\mathbb{R}^m : x \in \mathbb{S}, 0 \le t \le T\}$ is a $\mathcal{Z}$-adapted solution of the BSDE for a prescribed $\mathcal{Z}_T$-measurable terminal condition $Y_T = \gamma_T = \frac{\mathrm{d}\pi_T^\mu}{\mathrm{d}\pi_T^\nu}$.

For the HMM $(\mathcal{A},h)$, the relationship between the BSDE (8) and the backward map (1) is given of the following proposition:

*Proposition 3 (**Answer to Q2 in Sec. I**):* Fix T. Suppose Y_0 is the solution of (8) at time $t = 0$ and y_0 is defined according to the backward map (1). Then

$$Y_0(x) = y_0(x), \quad x \in \mathbb{S}$$

and along the solution (Y,V) of (8),

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{var}^\nu(Y_t(X_t)) = \mathsf{E}^\nu\big(\pi_t^\nu(\Gamma Y_t) + \pi_t^\nu(|V_t|^2)\big), \quad 0 \le t \le T \quad (9)$$

where $\mathrm{var}^\nu(Y_t(X_t)) := \mathsf{E}^\nu(|Y_t(X_t) - 1|^2)$ and $\pi_t^\nu(|V_t|^2) := \int_{\mathbb{S}} V_t^{\mathrm{T}}(x)V_t(x)\,\mathrm{d}\pi_t^\nu(x)$. (Eq. (9) is an example of a dissipation equation, and therefore an answer to question Q2 in Sec. I.)

Proof: See Appendix C. ■

Remark 3 (Relationship to (5)*):* Based on the dissipation equation (9) in Prop. 3, in order to obtain variance decay, a natural assumption is as follows:

(Assmp.) $\mathsf{E}^\nu\big(\pi_t^\nu(\Gamma Y_t) + \pi_t^\nu(|V_t|^2)\big) \ge c\,\mathrm{var}^\nu(Y_t(X_t)),\ 0 \le t \le T$

From (9) then $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{var}^\nu(Y_t(X_t)) \ge c\,\mathrm{var}^\nu(Y_t(X_t))$. Because $Y_T(X_T) = \gamma_T(X_T)$ at the terminal time, Gronwall implies $\mathrm{var}^\nu(Y_0(X_0)) \le e^{-cT}\mathrm{var}^\nu(\gamma_T(X_T))$ from which filter stability follows (see Prop. 2). In a prior conference paper [10], conditional Poincaré inequality (c-PI) is introduced for which the assumption above can be verified for $\nu = \bar{\mu}$. Several examples of Markov processes are described for which the c-PI holds. An important such example is the case where the Markov process satisfies the Doeblin condition.

Our aim in the remainder of this paper is to define appropriate notions of energy, variance, and Poincaré Inequality (PI) for HMM (Defn. 3) and relate the PI to the model properties (Sec. VI).

V. Poincaré Inequality (PI) for HMM

A. Function spaces and notation

Let $\rho \in \mathcal{P}(\mathbb{S})$ and $\tau > 0$. These are used to denote a generic prior and a generic time-horizon $[0,\tau]$. (In the analysis of filter stability, these are fixed to $\rho = \nu$ and $\tau = T$). The space of Borel-measurable deterministic functions is denoted

$$L^2(\rho) = \{f : \mathbb{S} \to \mathbb{R} : \rho(f^2) = \int_{\mathbb{S}} |f(x)|^2\,\mathrm{d}\rho(x) < \infty\}$$

Background from nonlinear filtering: A standard approach is based upon the Girsanov change of measure. Because the Novikov's condition holds, define a new measure $\tilde{\mathsf{P}}^\rho$ on $(\Omega,\mathcal{F}_\tau)$ as follows:

$$\frac{\mathrm{d}\tilde{\mathsf{P}}^\rho}{\mathrm{d}\mathsf{P}^\rho} = \exp\Big(-\int_0^\tau h^{\mathrm{T}}(X_t)\,\mathrm{d}W_t - \tfrac{1}{2}\int_0^\tau |h(X_t)|^2\,\mathrm{d}t\Big) =: D_\tau^{-1}$$

Then the probability law for X is unchanged but Z is a $\tilde{\mathsf{P}}^\rho$-B.M. that is independent of X [11, Lem. 1.1.5]. The expectation with respect to $\tilde{\mathsf{P}}^\rho$ is denoted by $\tilde{\mathsf{E}}^\rho(\cdot)$. The unnormalized filter $\sigma_\tau^\rho(f) := \tilde{\mathsf{E}}^\rho(D_\tau f(X_\tau)|\mathcal{Z}_\tau)$ for $f \in C_b(\mathbb{S})$. It is called as such because $\pi_\tau^\rho(f) = \frac{\sigma_\tau^\rho(f)}{\sigma_\tau^\rho(1)}$.

In a prior work, we introduced a dual optimal control formulation of the nonlinear filter [12], [3]. This requires consideration of the following Hilbert spaces:

• **Hilbert space for the dual:** Formally, the "dual" is a function on the state-space $\mathbb{S}$. The space of such functions is denoted as $\mathcal{Y}$. For the case when $\mathbb{S} = \{1,2,\cdots,d\}$, $\mathcal{Y} = \mathbb{R}^d$. Related to the dual, two types of Hilbert spaces are of interest. These are defined as follows:

- Hilbert space of $\mathcal{Z}_\tau$-measurable random functions:
  $$\mathbb{H}_\tau^\rho := \{F : \Omega \to \mathcal{Y} : F \in \mathcal{Z}_\tau \ \&\ \tilde{\mathsf{E}}^\rho(\sigma_\tau^\rho(F^2)) < \infty\}$$
  (This function space is important because the backward map (1) is a map from $\gamma_T \in \mathbb{H}_T^\nu$ to $y_0 \in L^2(\nu)$).
- Hilbert space of $\mathcal{Y}$-valued $\mathcal{Z}$-adapted stochastic processes:
  $$\mathbb{H}^\rho([0,\tau]) := \{Y : \Omega\times[0,\tau] \to \mathcal{Y} : Y_t \in \mathcal{Z}_t,\ 0 \le t \le \tau, \ \&\ \tilde{\mathsf{E}}^\rho\Big(\int_0^\tau \sigma_t^\rho(Y_t^2)\,\mathrm{d}t\Big) < \infty\}$$
  (This function space is important because the solution Y of the BSDE is an element of $\mathbb{H}^\rho([0,\tau])$).

Notation: Let $\rho \in \mathcal{P}(\mathbb{S})$. For real-valued functions $f,g \in \mathcal{Y}$, $\mathcal{V}_t^\rho(f,g) := \pi_t^\rho\big((f - \pi_t^\rho(f))(g - \pi_t^\rho(g))\big)$. With $f = g$, $\mathcal{V}_t^\rho(f) := \mathcal{V}_t^\rho(f,f)$.

B. Definitions of energy, variance, and PI

Dual optimal control system:

$$\begin{aligned} -\mathrm{d}Y_t(x) &= \big((\mathcal{A}Y_t)(x) - h^{\mathrm{T}}(x)\mathcal{V}_t^\rho(h,Y_t) \\ &\quad + h^{\mathrm{T}}(x)(V_t(x) - \pi_t^\rho(V_t))\big)\,\mathrm{d}t - V_t^{\mathrm{T}}(x)\,\mathrm{d}Z_t,\ 0 \le t \le \tau \\ Y_\tau(x) &= F(x),\ x \in \mathbb{S} \end{aligned} \qquad (10)$$

$(Y,V) \in \mathbb{H}^\rho([0,\tau])\times\mathbb{H}^\rho([0,\tau])^m$ is the solution of (10) for a given $F \in \mathbb{H}_\tau^\rho$. The dual optimal control system is important because of the following relationship to the nonlinear filter:

Proposition 4 (Prop. 1 in [3]): Consider (10). Then for a.e. $t \in [0,\tau]$,

$$\pi_t^\rho(Y_t) = \rho(Y_0) + \int_0^t \left(\mathcal{V}_s^\rho(h,Y_s) + \pi_s^\rho(V_s)\right)^{\mathrm{T}} \mathrm{d}Z_s, \quad \mathsf{P}^\rho - \text{a.s.} \tag{11a}$$

$$\mathsf{E}^\rho\left(\mathcal{V}_t^\rho(Y_t)\right) = \tag{11b}$$

$$\operatorname{var}^\rho(Y_0(X_0)) + \mathsf{E}^\rho\left(\int_0^t \pi_s^\rho(\Gamma Y_s) + |\mathcal{V}_s^\rho(h,Y_s)|^2 + \mathcal{V}_s^\rho(V_s)\,\mathrm{d}s\right)$$

Remark 4: The BSDE embedding (8) of the backward map (1) is a special case of (10). In particular, with $F = \gamma_T$, using (11a) with $\tau = T$ and $\rho = \nu$,

$$\mathcal{V}_t^\nu(h,Y_t) + \pi_t^\nu(V_t) = 0, \quad \mathsf{P}^\nu\text{-a.s., } 0 \le t \le T$$

(because $\pi_T^\nu(\gamma_T) = 1$). Therefore, (10) reduces to (8).

Let $\mathcal{N} := \{\rho \in \mathcal{P}(\mathbb{S}) : \operatorname{var}^\rho(Y_0(X_0)) = 0 \ \forall F \in \mathbb{H}_\tau^\rho\}$.

Definition 3: Consider (10). *Energy* is defined as follows:

$$\mathcal{E}^\rho(F) := \mathsf{E}^\rho\left(\int_0^\tau \pi_t^\rho(\Gamma Y_t) + |\mathcal{V}_t^\rho(h,Y_t)|^2 + \mathcal{V}_t^\rho(V_t)\,\mathrm{d}t\right)$$

For $\rho \in \mathcal{P}(\mathbb{S}) \setminus \mathcal{N}$, consider

$$\beta^\rho := \inf\left\{\mathcal{E}^\rho(F) \,:\, F \in \mathbb{H}_\tau^\rho \ \& \ \operatorname{var}^\rho(Y_0(X_0)) = 1\right\}$$

and the *Poincaré constant* is defined as follows:

$$c^\rho := \begin{cases} \dfrac{1}{\tau}\log\left(1+\beta^\rho\right), & \rho \in \mathcal{P}(\mathbb{S}) \setminus \mathcal{N} \\ 0, & \rho \in \mathcal{N} \end{cases}$$

Remark 5: The reason for defining the Poincaré constant in this manner is that c^ρ then represents a rate. In particular, using (11b), for each $\rho \in \mathcal{P}(\mathbb{S}) \setminus \mathcal{N}$,

$$\operatorname{var}^\rho(Y_0(X_0)) \le e^{-\tau c^\rho}\,\mathsf{E}^\rho\left(\mathcal{V}_\tau^\rho(F)\right), \quad \forall F \in \mathbb{H}_\tau^\rho$$

C. Existence of minimizer

We are interested in existence of the minimizers of the energy functional $\mathcal{E}^\rho(F)$ for $F \in \mathbb{H}_\tau^\rho$. If it exists, a minimizer is not unique because of the following translation symmetry:

$$\mathcal{E}^\rho(F + \alpha 1) = \mathcal{E}^\rho(F)$$

for any $\mathcal{Z}_\tau$-measurable random variable α such that $\tilde{\mathsf{E}}^\rho(\alpha^2) < \infty$. For this reason, consider the subspace

$$\mathcal{S}^\rho := \{F \in \mathbb{H}_\tau^\rho \,:\, \pi_\tau^\rho(F) = 0,\ \mathsf{P}^\rho - a.s.\}$$

Then $\mathcal{S}^\rho$ is closed subspace. (Suppose $F^{(n)} \to F$ in $\mathbb{H}_\tau^\rho$ with $\pi_\tau^\rho\left(F^{(n)}\right) = 0$. Then $\mathsf{E}^\rho\left(|\pi_\tau^\rho(F)|\right) = \mathsf{E}^\rho\left(|\pi_\tau^\rho(F - F^{(n)})|\right) \le \mathsf{E}^\rho\left(\pi_\tau^\rho(|F - F^{(n)}|^2)\right) = \tilde{\mathsf{E}}^\rho\left(\sigma_\tau^\rho(|F - F^{(n)}|^2)\right) = \|F - F^{(n)}\|_{\mathbb{H}_\tau^\rho} \to 0$.).

Proposition 5: Consider the optimal control system (10) with $Y_T = F \in \mathcal{S}^\rho$. Then $\rho(Y_0) = 0$ and

$$\mathcal{V}_t^\rho(h,Y_t) + \pi_t^\rho(V_t) = 0, \quad \mathsf{P}^\rho\text{-a.s., } 0 \le t \le \tau$$

Proof: The result follows from using (11a) in Prop. 4 (similar to Rem. 4). ■

Because of Prop. 5, for $Y_T = F \in \mathcal{S}^\rho$, (10) simplifies to

$$\begin{aligned} -\mathrm{d}Y_t(x) &= \left((\mathcal{A}Y_t)(x) + h^{\mathrm{T}}(x)V_t(x)\right)\mathrm{d}t - V_t^{\mathrm{T}}(x)\,\mathrm{d}Z_t,\ 0 \le t \le \tau \\ Y_\tau &= F \in \mathcal{S}^\rho,\ x \in \mathbb{S} \end{aligned} \tag{12}$$

Note that this is identical to the BSDE embedding (8) of the backward map. Its solution is used to define a linear operator as follows:

$$\mathsf{L}_0 : \mathcal{S}^\rho \subset \mathbb{H}_\tau^\rho \to L^2(\rho) \quad \text{by} \quad \mathsf{L}_0(F) := Y_0$$

(It is noted that (12) and therefore L_0 do not depend upon ρ even though the optimal control system (10) does). Additional details concerning this operator appear in Appendix D where it is shown that L_0 is bounded with $\|\mathsf{L}_0\| \le 1$.

The following Lemma provides sufficient condition for a minimizer to exist:

Lemma 2: Let $\rho \in \mathcal{P}(\mathbb{S}) \setminus \mathcal{N}$. Suppose that L_0 is compact. Then there exists an $F \in \mathcal{S}^\rho$ such that

$$\beta^\rho = \mathcal{E}^\rho(F) \quad \text{and} \quad \operatorname{var}^\rho(Y_0(X_0)) = 1$$

Proof: See Appendix D. ■

VI. PI and HMM model properties

In this section, we make the following assumption:

Assumption 2: The state-space is finite:

$$\textbf{(A2)} \qquad \mathbb{S} = \{1,2,\cdots,d\}$$

Notation: The space of functions and measures are both identified with $\mathbb{R}^d$: a real-valued function f (resp., a measure μ) is identified with a column vector in $\mathbb{R}^d$ where the x^{th} element of the vector equals $f(x)$ (resp., $\mu(x)$) for $x \in \mathbb{S}$, and $\mu(f) = \mu^{\mathrm{T}} f$. In this manner, the observation function $h : \mathbb{S} \to \mathbb{R}^m$ is also identified with a matrix $H \in \mathbb{R}^{d\times m}$. Its j-th column is denoted H^j for $j = 1,2,\cdots,m$. The constant function $1 = [1,1,\cdots,1]$ is a d-dimensional vector with all entries equal to one. $\mathcal{P}(\mathbb{S})$ is the probability simplex in $\mathbb{R}^d$. The generator $\mathcal{A}$ is identified with a transition rate matrix, denoted as A, whose (i,j) entry (for $i \ne j$) gives the non-negative rate of transition from state $i \mapsto j$. The diagonal entry (i,i) is chosen such that the sum of the elements in the i-th row is zero. The finite state-space HMM is denoted as (A,H). For any function $g : \mathbb{S} \to \mathbb{R}$, the notation

$$hg := \{\operatorname{diag}(H^j)g : j = 1,2,\cdots,m\}$$

For $m = 1$, this is simply the element-wise multiplication of the function h and g ($hg(x) = h(x)g(x)$ for $x \in \mathbb{S}$).

Definition 4: Consider an HMM (A,H) on a finite state-space $\mathbb{S} = \{1,2,\cdots,d\}$. The space of *observable functions* is the smallest subspace $\mathcal{O} \subset \mathbb{R}^d$ that satisfies the following two properties:

1) The constant function $1 \in \mathcal{O}$; and
2) If $g \in \mathcal{O}$ then $Ag \in \mathcal{O}$ and $hg \in \mathcal{O}$.

The space of *null eigenfunctions* is defined as

$$S_0 := \{f \in \mathbb{R}^d \mid Af = 0\}$$

Definition 5: Consider an HMM (A,H) on a finite state-space $\mathbb{S} = \{1,2,\cdots,d\}$.

1) The HMM is *observable* if $\mathcal{O} = \mathbb{R}^d$.
2) The Markov process is *ergodic* if
$$Af = 0 \implies f = c1$$
3) The HMM is *detectable* if $S_0 \subset \mathcal{O}$.

Remark 6: For additional motivation and background for these definitions, see [2, Sec. IV] and [4, Ch. 8]. It is shown (see [2, Rem. 13]) that the definition is equivalent to the (standard) definition of observability and detectability of HMM introduced in [6].

Example 1: Consider an HMM on $\mathbb{S} = \{1,2\}$ with

$$A = \begin{bmatrix} -\lambda_{12} & \lambda_{12} \\ \lambda_{21} & -\lambda_{21} \end{bmatrix}, \quad H = \begin{bmatrix} h(1) \\ h(2) \end{bmatrix}$$

For this model, the carré du champ operator and the observable space are as follows:

$$\Gamma f = \begin{bmatrix} \lambda_{12} \\ \lambda_{21} \end{bmatrix} (f(1) - f(2))^2, \quad \mathcal{O} = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} h(1) \\ h(2) \end{bmatrix} \right\}$$

Consequently,

1) A is ergodic iff $(\lambda_{12} + \lambda_{21}) > 0$. In this case, the invariant measure $\bar{\mu} = \left[\frac{\lambda_{21}}{(\lambda_{12}+\lambda_{21})} \quad \frac{\lambda_{12}}{(\lambda_{12}+\lambda_{21})} \right]^{\mathsf{T}}$.
2) (A,H) is is observable iff $h(1) \neq h(2)$.

A. Main result

*Proposition 6 (**Answer to Q3 in Sec. I**):* Consider the HMM (A,H) on finite state-space and $\rho \in \mathcal{P}(\mathbb{S}) \setminus \mathcal{N}$. Suppose any one of the following conditions holds:

1) The Markov process is ergodic; or
2) The HMM is observable; or
3) The HMM is detectable.

Then $c^\rho > 0$.

Proof: See Appendix E. ∎

The converse of this result – which gives the tightest condition for c^ρ to be positive – is as follows:

*Proposition 7 (**Answer to Q3 in Sec. I**):* Consider the HMM (A,H) on finite state-space. If $c^\rho > 0$ for all $\rho \in \mathcal{P}(\mathbb{S}) \setminus \mathcal{N}$, then the HMM is detectable.

Proof: See Appendix F. ∎

VII. Numerical Example

Consider an HMM on $\mathbb{S} = \{1,2,3,4\}$ with the transition rate matrix given by

$$A(\varepsilon) = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 2 & -2 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 2 & -2 \end{pmatrix} + \varepsilon \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

There are two cases:

1) **Case 1:** $\varepsilon = 0$. The Markov process is not ergodic. The space of null eigenfunctions is given by,
$$S_0 = N(A) = \text{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \right\}$$
This shows that the subsets $\{1,2\}$ and $\{3,4\}$ are the two communicating classes.
2) **Case 2:** $\varepsilon > 0$. The Markov process is ergodic with only a single communicating class given by $\mathbb{S}$.

Consider three choices for the observation function:

$$h^1 = \begin{pmatrix} 2 \\ 0 \\ 2 \\ 0 \end{pmatrix}, \quad h^2 = \begin{pmatrix} 2 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad h^3 = \begin{pmatrix} 2 \\ 0 \\ -2 \\ 0 \end{pmatrix}$$

With $\varepsilon > 0$, the HMM is detectable for any of these choices. Therefore, the interesting case arises when $\varepsilon = 0$. In this case, the following sub-cases arise:

1) **Case 1.1:** $\varepsilon = 0$ and $H = h^1$. The system (A,H) is not detectable.
2) **Case 1.2:** $\varepsilon = 0$ and $H = h^2$. The system (A,H) is not observable, but it is detectable.
3) **Case 1.3:** $\varepsilon = 0$ and $H = h^3$. The system (A,H) is observable.

Fig. 1 depicts the expected value of the χ^2-divergence between π_t^μ and π_t^ν from $\mu = [0.25, 0.40, 0.30, 0.05]$ and $\nu = [0.1, 0.2, 0.3, 0.4]$. The parameters together with the estimated rate of convergence are summarized in Table VII. An Euler discretization with step-size 0.005 is used to simulate the HMM and the nonlinear filter. The expectation is approximated by averaging over 500 Monte-Carlo simulations.

TABLE I
Combinations of Simulation Parameters

ε	h	Model property	Rate of conv.
0	h^1	Not detectable	0
0	h^2	Non-ergodic but detectable	0.075
0	h^3	Observable	0.155
0.1	h^1	Ergodic with $h(1) = h(3)$	0.196
0.1	h^3	Ergodic with $h(1) \neq h(3)$	0.412

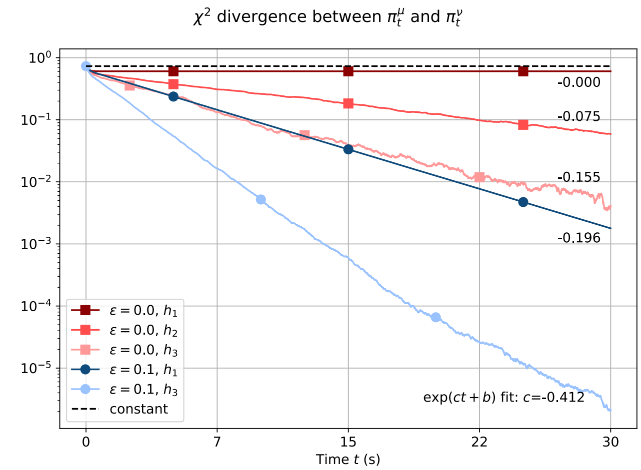

Fig. 1. $\mathsf{E}(\chi^2(\pi_t^\mu \mid \pi_t^\nu))$ for the model with different values of ε and observation function. The number on each line shows the exponential rate obtained from linear fitting.

REFERENCES

[1] D. Bakry, I. Gentil, and M. Ledoux, *Analysis and geometry of Markov diffusion operators*. Springer Science & Business Media, 2013, vol. 348.

[2] J. W. Kim and P. G. Mehta, "Duality for nonlinear filtering I: Observability," *IEEE Transactions on Automatic Control*, vol. 69, no. 2, pp. 699–711, 2024.

[3] ——, "Duality for nonlinear filtering II: Optimal control," *IEEE Transactions on Automatic Control*, vol. 69, no. 2, pp. 712–725, 2024.

[4] J. W. Kim, "Duality for nonlinear filtiering," Ph.D. dissertation, University of Illinois at Urbana-Champaign, Urbana, 06 2022.

[5] P. Chigansky, R. Liptser, and R. Van Handel, "Intrinsic methods in filter stability," in *Handbook of Nonlinear Filtering*, D. Crisan and B. Rozovskii, Eds. Oxford University Press, 2009.

[6] R. van Handel, "Observability and nonlinear filtering," *Probability Theory and Related Fields*, vol. 145, no. 1-2, pp. 35–74, 2009.

[7] J. M. C. Clark, D. L. Ocone, and C. Coumarbatch, "Relative entropy and error bounds for filtering of Markov processes," *Mathematics of Control, Signals and Systems*, vol. 12, no. 4, pp. 346–360, 1999.

[8] R. van Handel, "Nonlinear filtering and systems theory," in *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems*, 2010.

[9] A. Budhiraja, "Asymptotic stability, ergodicity and other asymptotic properties of the nonlinear filter," in *Annales de l'IHP Probabilités et Statistiques*, vol. 39, no. 6, 2003, pp. 919–941.

[10] J. W. Kim, P. G. Mehta, and S. Meyn, "The conditional Poincaré inequality for filter stability," in *2021 IEEE 60th Conference on Decision and Control (CDC)*, 12 2021, pp. 1629–1636.

[11] R. van Handel, "Filtering, stability, and robustness," Ph.D. dissertation, California Institute of Technology, Pasadena, 12 2006.

[12] J. W. Kim, P. G. Mehta, and S. Meyn, "What is the Lagrangian for nonlinear filtering?" in *2019 IEEE 58th Conference on Decision and Control (CDC)*. Nice, France: IEEE, 12 2019, pp. 1607–1614.

[13] J. W. Kim and P. G. Mehta, "Variance decay property for filter stability," *IEEE Transactions on Automatic Control*, To appear. [Online]. Available: https://doi.org/10.1109/TAC.2024.3413573

APPENDIX

A. Dissipation equation for a Markov process

The semigroup $\{P_t : t \geq 0\}$ is a solution of the Kolmogorov equation, $\frac{\partial}{\partial t}(P_t f) = \mathcal{A}(P_t f)$ for $t \geq 0$. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{V}^{\bar{\mu}}(P_t f) = 2\bar{\mu}\big((P_t f)(\mathcal{A}(P_t f))\big) - 2\bar{\mu}(P_t f)\bar{\mu}\big(\mathcal{A}(P_t f)\big) = -\bar{\mu}(\Gamma(P_t f))$$

where $\bar{\mu}\big(\mathcal{A}(P_t f)\big) = 0$ because $\bar{\mu}$ is an invariant measure.

B. Proof of Proposition 2

Using inequalities (2) and (5),

$$|\mathsf{E}^{\mu}\big(\chi^2(\pi_T^{\mu} \mid \pi_T^{\nu})\big)|^2 \leq e^{-cT}\mathrm{var}^{\nu}\big(\gamma_T(X_T)\big)\chi^2(\mu|\nu)$$

Since $\mathrm{var}^{\nu}\big(\gamma_T(X_T)\big) = \mathsf{E}^{\nu}\big(|\gamma_T(X_T)-1|^2\big) = \mathsf{E}^{\nu}\big(\chi^2(\pi_T^{\mu} \mid \pi_T^{\nu})\big)$,

$$\mathsf{E}^{\mu}\big(\chi^2(\pi_T^{\mu} \mid \pi_T^{\nu})\big) \leq \frac{1}{R_T}e^{-cT}\chi^2(\mu|\nu)$$

where $R_T := \frac{\mathsf{E}^{\mu}\left(\chi^2(\pi_T^{\mu}|\pi_T^{\nu})\right)}{\mathsf{E}^{\nu}\left(\chi^2(\pi_T^{\mu}|\pi_T^{\nu})\right)} \geq \mathrm{essinf}_{x\in\mathbb{S}}\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x) = \underline{a}$.

C. Proof of Proposition 3

Apply Itô formula on $Y_t(X_t)$ to obtain

$$\mathrm{d}Y_t(X_t) = V_t^{\mathrm{T}}(X_t)\,\mathrm{d}W_t + \mathrm{d}N_t$$

where $\{N_t : t \geq 0\}$ is a martingale [3, Remark 1]. Integrating,

$$\gamma_T(X_T) = Y_t(X_t) + \int_t^T V_s^{\mathrm{T}}(X_s)\,\mathrm{d}W_s + \mathrm{d}N_t \qquad (13)$$

and therefore

$$Y_t(x) = \mathsf{E}^{\nu}\big(\gamma_T(X_T) \mid \mathcal{Z}_t \vee [X_t = x]\big), \quad x \in \mathbb{S}$$

In particular at time $t = 0$, we have $Y_0(x) = y_0(x)$. The variance of $Y_t(X_t)$ is also obtained from (13):

$$\begin{aligned}\mathsf{E}^{\nu}&\big(|\gamma_T(X_T)-1|^2\big)\\ &= \mathsf{E}^{\nu}\Big(|Y_t(X_t)-1|^2 + \int_t^T |V_s(X_s)|^2 + (\Gamma Y_s)(X_s)\,\mathrm{d}s\Big)\\ &= \mathrm{var}^{\nu}(Y_t(X_t)) + \mathsf{E}^{\nu}\Big(\int_t^T \pi_s^{\nu}(\Gamma Y_s) + \pi_s^{\nu}(|V_s|^2)\,\mathrm{d}s\Big)\end{aligned}$$

Upon differentiating both sides with t gives (9).

D. Proof of Lemma 2

Discussions on the map L_0 and $\mathcal{E}^{\rho}$: For $F \in \mathbb{H}_\tau^{\rho}$, note that

$$\|F\|_{\mathbb{H}_\tau^{\rho}}^2 = \tilde{\mathsf{E}}^{\rho}\big(\sigma_\tau^{\rho}(F^2)\big) = \mathsf{E}^{\rho}\big(\pi_\tau^{\rho}(F^2)\big) = \mathsf{E}^{\rho}\big(F(X_\tau)^2\big)$$

For $F \in \mathcal{S}^{\rho}$, $\pi_\tau(F) = 0$ and $\therefore$, $\mathsf{E}^{\rho}\big((F(X_\tau))^2\big) = \mathsf{E}^{\rho}\big(\mathcal{V}_\tau^{\rho}(F)\big)$. Eq. (11b) in Prop. 4 is thus expressed as

$$\|F\|_{\mathbb{H}_\tau^{\rho}}^2 = \rho(Y_0^2) + \mathcal{E}^{\rho}(F), \quad F \in \mathcal{S}^{\rho} \qquad (14)$$

This shows that $\mathsf{L}_0 : \mathbb{H}_\tau^{\rho} \to L^2(\rho)$ is bounded with $\|\mathsf{L}_0\| \leq 1$.

To obtain the minimizer, setting $(\tilde{Y}, \tilde{V})$ to be the solution to (12) with $Y_\tau = \tilde{F} \in \mathcal{S}^{\rho}$, the functional derivative is evaluated as follows:

$$\langle \nabla\mathcal{E}^{\rho}(F), \tilde{F}\rangle := 2\,\mathsf{E}^{\rho}\left(\int_0^\tau \pi_t^{\rho}(\Gamma(Y_t, \tilde{Y}_t) + \pi_t^{\rho}(V_t^{\mathrm{T}}\tilde{V}_t)\,\mathrm{d}t\right)$$

where note $\pi_t^{\rho}(V_t^{\mathrm{T}}\tilde{V}_t) := \int_{\mathbb{S}} V_t^{\mathrm{T}}(x)\tilde{V}_t(x)\,\mathrm{d}\pi_t^{\rho}(x)$. Using Cauchy-Schwarz and (14),

$$|\langle \nabla\mathcal{E}^{\rho}(F), \tilde{F}\rangle|^2 \leq 4\|F\|_{\mathbb{H}_\tau^{\rho}}^2\,\|\tilde{F}\|_{\mathbb{H}_\tau^{\rho}}^2 \qquad (15)$$

This shows that $\tilde{F} \mapsto \langle \nabla\mathcal{E}^{\rho}(F), \tilde{F}\rangle$ is a bounded linear functional as a map from $\mathcal{S}^{\rho} \subset \mathbb{H}_\tau^{\rho}$ into $\mathbb{R}$. With these formalities completed, we show a minimizer exists.

Proof of Lemma 2: Let β^{ρ} be the infimum. Consider a sequence $\{F^{(n)} \in \mathcal{S}^{\rho} : n = 1, 2, \ldots\}$ such that $\mathcal{E}^{\rho}(F^{(n)}) \downarrow \beta^{\rho}$ and $\rho\big((Y_0^{(n)})^2\big) = 1$ for each n, with $Y_0^{(n)} := \mathsf{L}_0(F^{(n)})$. Using (14),

$$\|F^{(n)}\|_{\mathbb{H}_\tau^{\rho}}^2 = 1 + \mathcal{E}^{\rho}(F^{(n)}) < C, \quad n = 1, 2, \ldots$$

Therefore, $F^{(n)}$ is a bounded sequence in the Hilbert space $\mathbb{H}_\tau^{\rho}$, and there exists a weak limit $F \in \mathbb{H}_\tau^{\rho}$ such that $F^{(n)} \rightharpoonup F$. Moreover, $F \in \mathcal{S}^{\rho}$ because $\mathcal{S}^{\rho}$ is closed. Therefore $Y_0 := \mathsf{L}_0(F)$ satisfies $\rho(Y_0) = 0$. Since $F^{(n)} \rightharpoonup F$ in $\mathbb{H}_\tau^{\rho}$ and L_0 is compact, we have $Y_0^{(n)} \to Y_0$ in $L^2(\rho)$. Therefore, $\rho\big((Y_0)^2\big) = \lim_{n\to\infty}\rho\big((Y_0^{(n)})^2\big) = 1$.

It remains to show that $\mathcal{E}^{\rho}(F) = \beta^{\rho}$. The map $F \mapsto \mathcal{E}^{\rho}(F)$ is convex. Therefore,

$$\mathcal{E}^{\rho}(F^{(n)}) \geq \mathcal{E}^{\rho}(F) + \langle \nabla\mathcal{E}^{\rho}(F), F^{(n)} - F\rangle$$

We have already shown that $\tilde{F} \mapsto \langle \nabla\mathcal{E}^{\rho}(F), \tilde{F}\rangle$ is a bounded linear functional. Therefore, letting $n \to \infty$, the second term on the right-hand side converges to zero and

$$\mathcal{E}^{\rho}(F) \leq \lim_{n\to\infty}\mathcal{E}^{\rho}(F^{(n)}) = \beta^{\rho}$$

Because β^{ρ} is the infimum, this show that $\mathcal{E}^{\rho}(F) = \beta^{\rho}$.

E. Proof of Proposition 6

Proof: [of Prop. 6] Suppose any one of the three conditions hold. We claim then

$$\text{(claim)} \quad \mathcal{E}^\rho(F) = 0 \implies \text{var}^\rho(Y_0(X_0)) = 0$$

Supposing the claim is true, the proof is by contradiction. To see this suppose $\beta^\rho = 0$. By Lemma 2, there exists $\mathcal{E}^\rho(F) = 0$ such that $\text{var}^\rho(Y_0(X_0)) = 1$ which contradicts the claim. It remains to prove the claim. For the three cases, its proof is described in the remainder of this appendix. ■

1. Ergodic case: At time $t > 0$, let ρ_t denote the probability law of X_t (without conditioning). Then $\rho \ll \rho_t$ and $\text{supp}(\rho_t) =: \mathbb{S}'$ is identical for any $t > 0$. W.l.o.g., take $\mathbb{S}'$ as the new state-space and consider the Markov process on $\mathbb{S}'$. Suppose $\mathcal{E}^\rho(F) = 0$. Then

$$\mathsf{E}^\rho\Big(\int_0^\tau \pi_t^\rho\big(\Gamma Y_t\big)\,\mathrm{d}t\Big) = 0 \implies \pi_t^\rho\big(\Gamma Y_t\big) = 0$$

almost every $t \in [0,\tau]$, $\mathsf{P}^\rho|_{\mathcal{Z}_\tau}$-almost surely. For white noise observation model, $\text{supp}(\pi_t^\rho) =: \mathbb{S}'$ and therefore $\Gamma Y_t(x) = 0$ for all $x \in \mathbb{S}'$, and therefore $\Gamma Y_t = 0$ with probability 1. If the model is ergodic, this implies Y_t is a constant function, and therefore $\mathsf{E}^\rho\big(\mathcal{V}_t^\rho(Y_t)\big) = 0$. The proof of the claim is completed by noting $\text{var}^\rho(Y_0(X_0)) \le \mathsf{E}^\rho(\mathcal{V}_t^\rho(Y_t))$ from (11b).

2. Observable case: The proof is based on using the equation for conditional covariance $\mathcal{V}_t^\rho(f, Y_t)$ (see [13, Appdx. D] for its proof):

$$\begin{aligned}\mathrm{d}\mathcal{V}_t^\rho(f,Y_t) &= \Big(\pi_t^\rho\big(\Gamma(f,Y_t)\big) + \mathcal{V}_t^\rho(\mathcal{A}f,Y_t)\Big)\mathrm{d}t \\ &+ \Big(\mathcal{V}_t^\rho\big((f-\pi_t^\rho(f))(h-\pi_t^\rho(f)),Y_t\big) + \mathcal{V}_t^\rho(f,V_t)\Big)^{\mathsf{T}}\mathrm{d}I_t^\rho\end{aligned} \tag{16}$$

The equation is used to prove the following Lemma which is the key to prove the claim.

Lemma 3: Suppose $\mathcal{E}^\rho(F) = 0$. Then for each $f \in \mathcal{O}$,

$$\mathcal{V}_t^\rho(f,Y_t) = 0, \quad \mathsf{P}^\rho\text{-a.s., a.e. } 0 \le t \le \tau$$

Proof: From the defining relation for $\mathcal{E}^\rho(F)$,

$$\pi_t^\rho(\Gamma Y_t) = 0,\ \mathcal{V}_t^\rho(h,Y_t) = 0,\ \mathcal{V}_t^\rho(V_t) = 0,\ \mathsf{P}^\rho\text{-a.s.}$$

for a.e. $0 \le t \le \tau$. Using the Cauchy-Schwarz formula then for each $f \in C_b(\mathbb{S})$,

$$|\mathcal{V}_t^\rho(f,V_t)|^2 \le \mathcal{V}_t^\rho(f)\mathcal{V}_t^\rho(V_t) = 0 \quad \mathsf{P}^\rho\text{-a.s.}$$

Similarly, upon using the Cauchy-Schwarz formula [1, Eq.1.4.3] for carré du champ, $\pi_t^\rho(\Gamma(f,Y_t)) = 0$, P^ρ-a.s.. Based on these (16) simplifies to

$$\begin{aligned}&\mathrm{d}\mathcal{V}_t^\rho(f,Y_t) \\ &= \mathcal{V}_t^\rho(\mathcal{A}f,Y_t)\,\mathrm{d}t + \big(\mathcal{V}_t^\rho(hf,Y_t) - \pi_t^\rho(h)\mathcal{V}_t^\rho(f,Y_t)\big)^{\mathsf{T}}\mathrm{d}I_t^\rho\end{aligned}$$

Therefore, $\mathcal{V}_t^\rho(f,Y_t) = 0,\ \ 0 \le t \le \tau$

$$\implies \quad \mathcal{V}_t^\rho(\mathcal{A}f,Y_t) = 0,\ \mathcal{V}_t^\rho(hf,Y_t) = 0,\ \ 0 \le t \le \tau$$

Since $\mathcal{V}_t^\rho(1,Y_t) = 0$ for all $t \in [0,\tau]$, the result follows from Defn. 4 of the observable space $\mathcal{O}$. ■

Based on the result in Lemma 3, if $\mathcal{O} = \mathbb{R}^d$, we have $\mathsf{E}^\rho\big(\mathcal{V}_t^\rho(Y_t)\big) = 0$, and then the claim follows because $\text{var}^\rho(Y_0(X_0)) \le \mathsf{E}^\rho(\mathcal{V}_t^\rho(Y_t))$ from (11b).

3. Detectable case: As in the ergodic case, if $\mathcal{E}^\rho(F) = 0$ then $\Gamma Y_t(x) = 0$ for all $x \in \mathbb{S}'$, and therefore $Y_t \in S_0$. If the system $(\mathcal{A}, h)$ is detectable, then this implies $Y_t \in \mathcal{O}$ with probability 1. By Lemma 3, $\mathsf{E}^\rho\big(\mathcal{V}_t^\rho(Y_t)\big) = 0$ and the claim follows.

F. Proof of Proposition 7

Suppose HMM is not detectable. Our goal is to find $\rho \in \mathcal{P}(\mathbb{S}) \setminus \mathcal{N}$ and $F \in \mathbb{H}_\tau^\rho$ such that $\mathcal{E}^\rho(F) = 0$ and $\text{var}^\rho(Y_0(X_0)) = 1$. We begin with two claims:

1) **Claim 1.** There exists a $\rho \in \mathcal{P}(\mathbb{S})$ and $f \in S_0$ such that (a) $\rho(f^2) > 0$, and (b) $\rho(fg) = 0$ for all $g \in \mathcal{O}$.
2) **Claim 2.** For any such f and ρ (that satisfy the two conditions in claim 1), $\pi_t^\rho(fg) \equiv 0$, $0 \le t \le \tau$, $\forall\, g \in \mathcal{O}$.

Assuming these claims are true $\pi_\tau^\rho(f) = 0$ and therefore $f \in \mathcal{S}_0$. Because $f \in S_0$, $\mathcal{A}f = 0$ and therefore, $Y_t \equiv f$ and $V_t \equiv 0$ solves the BSDE (12), whose energy $\mathcal{E}^\rho(f) = 0$ and $\text{var}^\rho(Y_0(X_0)) = \rho(f^2) > 0$. It remains to prove the two claims.

Proof of claim 2: Since $f \in S_0$, $\mathcal{A}f = 0$, and consequently, $\mathcal{A}(fg) = f\mathcal{A}g$. From (7):

$$\mathrm{d}\pi_t^\rho(fg) = \pi_t^\rho\big(f\mathcal{A}g\big)\,\mathrm{d}t + \pi_t^\rho(ghf)^{\mathsf{T}}\,\mathrm{d}I_t,\ 0 \le t \le \tau$$

Because $\mathcal{A}g \in \mathcal{O}$ and $gh \in \mathcal{O}$ (Defn. 4) this shows claim 2. At time $t = 0$, $\pi_t^\rho\big(f\mathcal{A}g\big)\big|_{t=0} = \rho(f\mathcal{A}g) = 0$ and $\pi_t^\rho(ghf)\big|_{t=0} = \rho(fgh) = 0$. Therefore, $\pi_t^\rho(fg) \equiv 0$ is the equilibrium soln.

Proof of claim 1: Consider an ergodic partition $S = \cup_k S^{(k)}$ and the invariant measures $\bar{\mu}^{(k)}$ with support on $S^{(k)}$ for each k. W.l.o.g, upon re-ordering of indices if necessary, there exists an $f = 1_{S^{(1)}} - 1_{S^{(2)}}$ such that $f \notin \mathcal{O}$. Therefore, w.l.o.g, we may restrict ourselves to ergodic partition with exactly two components, $S = S^{(1)} \cup S^{(2)}$, with invariant measures $\bar{\mu}^{(1)}$ and $\bar{\mu}^{(2)}$ with support on $S^{(1)}$ and $S^{(2)}$, respectively. For this case, $\dim(N(\mathcal{A})) = 2$ and since 1 is contained in both $N(\mathcal{A})$ and in $\mathcal{O}$, and the fact that $N(\mathcal{A}) \notin \mathcal{O}$, $\dim(N(\mathcal{A}) \cap \mathcal{O}) = 1$. Consider the restriction (noting $\mathcal{O}$ is $\mathcal{A}$-invariant),

$$\mathcal{A}|_{\mathcal{O}} : \mathcal{O} \to \mathcal{A}(\mathcal{O}) \subset \mathcal{O}$$

Then $\dim(N(\mathcal{A}|_{\mathcal{O}})) = 1$. By rank-nullity, $\dim(\mathcal{A}(\mathcal{O})) = \dim(\mathcal{O}) - 1$. Consider a decomposition $\mathcal{O} = \mathcal{A}(\mathcal{O}) \oplus \text{span}\{\bar{o}\}$ for some $\bar{o} \in \mathcal{O}$. Use the decomposition to express any $g \in \mathcal{O}$ as $g = g_A + a\bar{o}$, where $a \in \mathbb{R}$. Now, $\bar{\mu}^{(1)}(1) = \bar{\mu}^{(2)}(1) = 1 \implies 1 \notin \mathcal{A}(\mathcal{O})$ (since, for all $g \in \mathcal{A}(\mathcal{O})$, $\bar{\mu}^{(1)}(g) = \bar{\mu}^{(2)}(g) = 0$). Because $1 \in \mathcal{O}$, a possible choice is to pick $\bar{o} = 1$.

We now pick $\rho \in \mathcal{P}(\mathbb{S})$ and $f \in S_0$ to show that $\rho(fg) = 0$ for all $g \in \mathcal{O}$. Set

$$\rho = \tfrac{1}{2}\bar{\mu}^{(1)} + \tfrac{1}{2}\bar{\mu}^{(2)}, \quad f = c_1 1_{S^{(1)}} + c_2 1_{S^{(2)}}$$

where the constants c_1 and c_2 need to be picked. Because $g_A \in R(\mathcal{A})$, $\rho(fg_A) = \frac{1}{2}(c_1\bar{\mu}^{(1)}(g_A) + c_2\bar{\mu}^{(2)}(g_A)) = 0$. Therefore,

$$\rho(fg) = \frac{a}{2}(c_1\bar{\mu}^{(1)}(\bar{o}) + c_2\bar{\mu}^{(2)}(\bar{o}))$$

Pick constants c_1 and c_2 to make the right-hand side zero. With $\bar{o} = 1$, $c_1 = 1$ and $c_2 = -1$ works for which $\rho(f) = 0$ and $\rho(f^2) = 1$.