SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 8, No. 2, pp. 539–572

Uncertainty Quantification for Markov Processes via Variational Principles and Functional Inequalities*

Jeremiah Birrell† and Luc Rey-Bellet†

Abstract. Information-theory based variational principles have proven effective at providing scalable uncertainty quantification (i.e., robustness) bounds for quantities of interest in the presence of nonparametric model-form uncertainty. In this work, we combine such variational formulas with functional inequalities (Poincaré, log-Sobolev, Liapunov functions) to derive explicit uncertainty quantification bounds for time-averaged observables, comparing a Markov process to a second (not necessarily Markov) process. These bounds are well behaved in the infinite-time limit and apply to steady-states of both discrete and continuous-time Markov processes.

Key words. uncertainty quantification, Markov process, relative entropy, Poincaré inequality, log-Sobolev inequality, Liapunov function, Bernstein inequality

AMS subject classifications. 47D07, 39B72, 60F10, 60J25

DOI. 10.1137/19M1237429

1. Introduction.

Information-theory based variational principles have proven effective at providing uncertainty quantification (i.e., robustness) bounds for quantities of interest in the presence of nonparametric model-form uncertainty [15, 23, 51, 32, 41, 24, 33, 10, 43, 31]. In the present work, we combine these tools with functional inequalities to obtain improved and explicit uncertainty quantification (UQ) bounds for both discrete and continuous-time Markov processes on general state spaces.

In our approach, we are given a baseline model, described by a probability measure P; this is the model one has "in hand" and that is amenable to analysis/simulation, but it may contain many sources of error and uncertainty. Perhaps it depends on parameters with uncertain values (obtained from experiment, Monte Carlo simulation, variational inference, etc.) or is obtained via some approximation procedure (dimension reduction, neglecting memory terms, asymptotic approximation, etc.). In short, any quantity of interest computed from P has (potentially) significant uncertainty associated with it. Mathematically we chose to express this uncertainty by considering a (nonparametric) family, $\mathcal{U}(P)$, of alternative models that we postulate contains the inaccessible "true" model.

Loosely stated, given some observable F, the UQ goal considered here is

$$\textbf{Bound the bias } E_{\widetilde{P}}[F]-E_P[F], \text{ where } \widetilde{P}\in\mathcal{U}_r(P). \tag{1.1}$$

*Received by the editors January 8, 2019; accepted for publication (in revised form) January 24, 2020; published electronically April 21, 2020.
https://doi.org/10.1137/19M1237429
Funding: The research of the second author was partially supported by the National Science Foundation (NSF) under grant DMS-1515712 and the Air Force Office of Scientific Research (AFOSR) under grant FA-9550-18-1-0214.
†Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA 01003 (birrell@math.umass.edu, luc@math.umass.edu).

The subscript r indicates that the "neighborhood" of alternative models, $\mathcal{U}_r(P)$, is often defined in terms of an error tolerance, $r > 0$. For our purposes, the appropriate notion of neighborhood will be expressed in terms of relative entropy, which can be interpreted as measuring the loss of information due to uncertainties. We do not discuss in full generality how to choose the tolerance level r, but there are cases where one has enough information about the "true" model to choose an appropriate tolerance; see section 6.

Remark 1.1. Note that in (1.1) and the remainder of this paper, we consider the case where the quantity of interest is the expected value of some function, but extensions of these ideas to other quantities of interest are possible [24].

More specifically, here we work with a Markov process (X_t, P^μ) with initial distribution μ and stationary distribution μ^*, and we compare it to an alternative (not necessarily Markov) process $(X_t, \widetilde{P}^{\widetilde{\mu}})$; we study the problem of bounding the bias when the finite-time averages of a real-valued observable, f, are computed by sampling from the alternative process:

$$\textbf{Bound} \qquad \widetilde{E}^{\widetilde{\mu}}\left[\frac{1}{T}\int_0^T f(X_t)dt\right] - \int f d\mu^*. \tag{1.2}$$

Here, $\widetilde{E}^{\widetilde{\mu}}$ denotes the expectation with respect to $\widetilde{P}^{\widetilde{\mu}}$, and similarly for P^μ, E^μ. (Discrete-time processes will also be considered in section 5.)

Equation (1.2) is a (less studied) variant of the classical problem of the convergence of ergodic averages to the expectation in the stationary distribution:

$$E^\mu\left[\frac{1}{T}\int_0^T f(X_t)dt\right] \to \int f d\mu^*. \tag{1.3}$$

By combining information on the problems (1.2) and (1.3), one can also obtain bounds on the finite-time sampling error:

$$\mathrm{err}_T = E^\mu\left[\frac{1}{T}\int_0^T f(X_t)dt\right] - \widetilde{E}^{\widetilde{\mu}}\left[\frac{1}{T}\int_0^T f(X_t)dt\right]. \tag{1.4}$$

In this work, we focus on the robustness problem, (1.2).

There are classical inequalities addressing (1.1) (for example, Csiszar–Kullback–Pinsker, Le Cam, Scheffé, etc.), but they exhibit poor scaling properties with problem dimension and/or in the infinite-time limit and so are inappropriate for bounding (1.2). This problem can be addressed by using tight information inequalities based on the Gibbs variational principle that are summarized in section 2. See [41] for a detailed discussion of these issues.

Other recent works have also focused on robustness bounds for Markov processes, often with the goal of providing error bounds for approximate Markov chain Monte Carlo samplers. Bounds on the difference between the distributions (finite-time or stationary) of Markov processes have been obtained in both total-variation [50, 28, 1, 49, 5, 39] and Wasserstein distances [54, 52, 37].

One benefit of the approach taken in the present work is that the bounds naturally incorporate information on the specific observable, f, under investigation; for instance, the asymptotic variance of f under the baseline model appears in the bound in Theorem 4.6 below. When the end goal is robustness bounds for time-averages of f, this observable specificity has the potential to yield tighter bounds; see also [52] for bounds that incorporate similar information on the observable.

Our method utilizes relative entropy to quantify the distance between models. A drawback, compared to the total-variation and Wasserstein distance approaches, is the requirement of absolute continuity; however, this is satisfied in many cases of interest. As we will see, one benefit of utilizing relative entropy is that the alternative model does not have to be a Markov process. The second main innovation here is the use of various functional inequalities, in combination with relative entropy, to bound (1.2). The end result is computable, finite-time UQ bounds that are also well behaved in the long-time limit.

1.1. Summary of results. The basis for all of our bounds is Theorem 2.11 in section 2,

$$\begin{aligned}(1.5)\qquad &\pm\left(\widetilde{E}^{\widetilde{\mu}}\left[\frac{1}{T}\int_0^T f(X_t)dt\right]-\mu^*[f]\right)\\ &\leq \inf_{c>0}\left\{\frac{1}{cT}\Lambda^{\widehat{f}_T}_{P^{\mu^*}_T}(\pm c)+\frac{1}{cT}R(\widetilde{P}^{\widetilde{\mu}}_T\|P^{\mu^*}_T)\right\},\quad \widehat{f}_T\equiv\int_0^T f(X_t)-\mu^*[f]dt,\end{aligned}$$

along with Corollary 3.5 in section 3,

$$(1.6)\qquad \frac{1}{T}\Lambda^{\widehat{f}_T}_{P^{\mu^*}_T}(\pm c)\leq\kappa(V_{\pm c}),$$

$$(1.7)\qquad \kappa(V)\equiv\sup\left\{\langle A[g],g\rangle+\int V|g|^2d\mu^*: g\in D(A,\mathbb{R}),\ \|g\|_{L^2(\mu^*)}=1\right\},$$

$$(1.8)\qquad V_{\pm c}(x)\equiv\pm c\left(f(x)-\mu^*[f]\right),\ \ \mu^*[f]\equiv\int f d\mu^*.$$

In the above, $\Lambda^{\widehat{f}_T}_{P^{\mu^*}_T}(\pm c)$ is the cumulant generating function of $\widehat{f}_T$ (see (2.24) for details), $R(\widetilde{P}^{\widetilde{\mu}}_T\|P^{\mu^*}_T)$ is the relative entropy of the processes up to time T (see (2.5)), $\langle\cdot,\cdot\rangle$ denotes the inner product on $L^2(\mu^*)$, and $(A,D(A,\mathbb{R}))$ is the generator of the Markov semigroup for the process (X_t,P^μ) on $L^2(\mu^*)$. Again, we emphasize that the alternative process, $(X_t,\widetilde{P}^{\widetilde{\mu}})$, does *not* need to be Markov; for an example involving semi-Markov processes, see section 6.2.

Equation (1.5) is derived by employing the Gibbs variational principle (hence the relation to relative entropy). Equation (1.6), which is based on a theorem proven in [56], results from a connection between the cumulant generating function and the Feynman–Kac semigroup (hence the appearance of the generator, A). Also, note that the bound is expected to behave well in the limit $T\to\infty$, as $R(\widetilde{P}^{\widetilde{\mu}}_T\|P^{\mu^*}_T)/T$ converges to the relative entropy rate of the processes, under suitable ergodicity assumptions.

Equation (1.6) allows us to employ our primary new tool for UQ, that is, functional inequalities. By functional inequalities, we mean bounds on the generator, A, that will yield bounds on $\kappa(V_{\pm c})$; we will cover Poincaré, log-Sobolev, and F-Sobolev inequalities, as well as Liapunov functions. Our results rely heavily on the bounds obtained in [56, 45, 12, 34, 29], where concentration inequalities for ergodic averages were obtained.

The method outlined above leads to explicit UQ bounds, expressed in terms of the following quantities:

1. Spectral properties of the generator, A, of the dynamics of the baseline model, P, in the stationary regime (i.e., on $L^2(\mu^*)$); see (1.5)–(1.7). This term depends on the chosen observable but does *not* depend on the alternative model; functional inequalities are only required for the base model (which is often the simpler model). This is one of the strengths of the method, though computing explicit, tight constants for these functional inequalities is still a very difficult problem in general.
2. The path-space relative entropy up to time T, $R(\widetilde{P}_T^{\widetilde{\mu}}||P_T^{\mu^*})$, of the alternative model with respect to the base. This term depends heavily on the difference in dynamics between the two models; in particular, a nontrivial bound requires absolute continuity of the path-space distributions. This is a drawback of the relative-entropy based method employed here, but it does hold in many cases of interest; see section 6 for examples.

While most of our results do not assume reversibility of the base process, bounds based on (1.7) only involve the symmetric part of the generator and so are generally less than ideal, or even useless, for many nonreversible systems. This is a drawback of the approach pursued here.

Remark 1.2. For certain hypocoercive systems, ergodicity can be proven by working with an alternative metric; see [20, 21, 16]. It is possible that the functional-inequality based UQ techniques developed below could be adapted to this more general setting; a step in that direction can be found in [6].

For a simple example of the type of result obtained below, consider diffusion on $\mathbb{R}^n$ in a C^2 potential, V; i.e., the generator is $A = \Delta - \nabla V \cdot V$, and the invariant measure is $\mu^* = e^{-V(x)}dx$. Suppose the Hessian of V is bounded below:

$$D^2 V(x) \geq \alpha^{-1} I, \quad \alpha > 0. \tag{1.9}$$

Our results give a Bernstein-type UQ bound for any bounded f:

$$\pm\left(\widetilde{E}^{\widetilde{\mu}}\left[\frac{1}{T}\int_0^T f(X_t)dt\right] - \mu^*[f]\right) \leq \sqrt{2\sigma^2\eta} + M^{\pm}\eta, \tag{1.10}$$
$$M^{\pm} = \alpha\|(f-\mu^*[f])^{\pm}\|_\infty, \;\; \sigma^2 = 2\alpha \operatorname{Var}_{\mu^*}[f], \;\; \eta = \frac{1}{T}R(\widetilde{P}_T^{\widetilde{\mu}}||P_T^{\mu^*}).$$

(This bound can also be improved by using the asymptotic variance; see section 4.2.) Section 4.4.1 contains further discussion of diffusions.

The remainder of the paper is structured as follows. Necessary background on UQ for both general measures and processes will be given in section 2, leading up to a connection with both the Feynman–Kac semigroup and the relative entropy rate. Relevant properties of the Feynman–Kac semigroup are given in section 3, culminating with the bound (1.6). The use of functional inequalities to obtain explicit UQ bounds from (1.6) will be explored in section 4. In section 5 we show how these ideas can be adapted to discrete-time processes. Finally, the problem of bounding the relative entropy rate will be addressed in section 6.

2. UQ for Markov processes.

2.1. UQ via variational principles. In this subsection, we recall several earlier results regarding the variational-principle approach to UQ, as developed in [15, 23, 10, 31, 8]. In particular, Proposition 2.2, quoted from [23], will be a critical tool in our derivation of UQ bounds for Markov processes.

Let P be a probability measure on a measurable space $(\Omega, \mathcal{F})$. We consider the class of random variables $f : \Omega \to \mathbb{R}$ with a well-defined and finite moment generating function in a neighborhood of the origin:

$$\mathcal{E}(P) = \left\{ f : \Omega \to \mathbb{R} \,:\, E_P[e^{\pm c_0 f}] < \infty \text{ for some } c_0 > 0 \right\}. \tag{2.1}$$

It is not difficult to prove (see, e.g., [17]) that the cumulant generating function

$$\Lambda_P^f(c) = \log E_P[e^{cf}] \tag{2.2}$$

is a convex function, finite and infinitely differentiable in some interval (c_-, c_+) with $-\infty \leq c_- < 0 < c_+ \leq \infty$ and equal to $+\infty$ outside of $[c_-, c_+]$. Moreover if $f \in \mathcal{E}(P)$, then f has moments of all orders, and we write

$$\widehat{f} = f - E_P[f] \tag{2.3}$$

for the centered observable of mean 0. We will often use the cumulant generating function for the centered observable $\widehat{f}$:

$$\Lambda_P^{\widehat{f}}(c) = \log E_P[e^{c(f - E_P[f])}] = \Lambda_P^f(c) - cE_P[f]\,. \tag{2.4}$$

Recall also the relative entropy (or Kullback–Leibler divergence), defined by

$$R(\widetilde{P}||P) = \begin{cases} E_{\widetilde{P}}\left[\log\left(\frac{d\widetilde{P}}{dP}\right)\right] & \text{if } \widetilde{P} \ll P, \\ +\infty & \text{otherwise.} \end{cases} \tag{2.5}$$

It has the property of a divergence; that is, $R(\widetilde{P}||P) \geq 0$ and $R(\widetilde{P}||P) = 0$ if and only if $\widetilde{P} = P$.

A key ingredient in our approach is the Gibbs variational principle, which relates the cumulant generating function and relative entropy; see Proposition 1.4.2 in [22].

Proposition 2.1 (Gibbs variational principle). *Let $f : \Omega \to \mathbb{R}$ be bounded and measurable. Then*

$$\log E_P[e^f] = \sup_{\widetilde{P}: R(\widetilde{P}||P) < \infty} \left\{ E_{\widetilde{P}}[f] - R(\widetilde{P}||P) \right\}. \tag{2.6}$$

As shown in [15, 23], the Gibbs variational principle implies the following UQ bounds for the expected values (a similar inequality is used in the context of concentration inequalities—see, e.g., [8]—and was also used independently in [10, 31]). For a proof of the version stated here, see pages 85–86 in [23].

Proposition 2.2 (Gibbs information inequality). *If $R(\widetilde{P}||P) < \infty$ and $f \in \mathcal{E}(P)$, then $f \in L^1(\widetilde{P})$ and*

$$-\inf_{c>0}\left\{\frac{\Lambda_P^{\widehat{f}}(-c)+R(\widetilde{P}||P)}{c}\right\} \leq E_{\widetilde{P}}[f]-E_P[f] \leq \inf_{c>0}\left\{\frac{\Lambda_P^{\widehat{f}}(c)+R(\widetilde{P}||P)}{c}\right\}. \tag{2.7}$$

Remark 2.3. Note that even if $R(\widetilde{P}||P)=\infty$, the bound (2.7) trivially holds as long as $E_{\widetilde{P}}[f]$ is defined. To avoid clutter in the statement of our results, when $R(\widetilde{P}||P)=\infty$ we will consider the bound to be satisfied for any $f \in \mathcal{E}(P)$, even if $E_{\widetilde{P}}[f]$ is undefined.

Optimization problems of the form in (2.7) will appear frequently; hence we write the following definition.

Definition 2.4. *Given any $\Lambda : \mathbb{R} \to [0,\infty]$ and $\eta > 0$, let*

$$\Xi^{\pm}(\Lambda,\eta) \equiv \inf_{c>0}\left\{\frac{\Lambda(\pm c)+\eta}{c}\right\}. \tag{2.8}$$

With this, we can rewrite the bound (2.7) as

$$-\,\Xi^{-}\left(\Lambda_P^{\widehat{f}}, R(\widetilde{P}||P)\right) \leq E_{\widetilde{P}}[f]-E_P[f] \leq \Xi^{+}\left(\Lambda_P^{\widehat{f}}, R(\widetilde{P}||P)\right). \tag{2.9}$$

Inequality (2.9) is the starting point for all UQ bounds derived in this paper. From it, we see which quantities must be controlled in order to make the UQ bounds explicit: the relative entropy and the cumulant generating function. The former will be discussed in section 6. For Markov processes, the latter can be bounded via a connection with the Feynman–Kac semigroup and functional inequalities; this connection between functional inequalities and UQ bounds is the main focus and innovation of the current work, and we begin discussing it in section 2.4. First we recall some general properties of the bounds (2.9).

2.2. Properties of $\Xi^{\pm}$.

The objects

$$\Xi(\widetilde{P}||P;\pm f) \equiv \Xi^{\pm}\left(\Lambda_P^{\widehat{f}}, R(\widetilde{P}||P)\right) \tag{2.10}$$

appearing in the Gibbs information inequality, (2.9), have many remarkable properties, of which we recall a few.

Proposition 2.5. *Assume $R(\widetilde{P}||P) < \infty$ and $f \in \mathcal{E}(P)$. We have the following:*

1. (Divergence) *$\Xi(\widetilde{P}||P;f)$ is a divergence, i.e., $\Xi(\widetilde{P}||P,f) \geq 0$ and $\Xi(\widetilde{P}||P;f)=0$ if and only if either $P=\widetilde{P}$ or f is constant P-a.s.*
2. (Linearization) *If $R(\widetilde{P}||P)$ is sufficiently small, we have*

$$\Xi(\widetilde{P}||P,f) = \sqrt{2\mathrm{Var}_P[f]R(\widetilde{P}||P)} + O(R(\widetilde{P}||P)). \tag{2.11}$$

3. (Tightness) *For $\eta > 0$ consider $\mathcal{U}_\eta = \{\widetilde{P}\,;\, R(\widetilde{P}||P) \leq \eta\}$. There exists η^* with $0 < \eta^* \leq \infty$ such that for any $\eta < \eta^*$ there exists a measure P^η with*

$$\sup_{\widetilde{P}\in\mathcal{U}_\eta}\left\{E_{\widetilde{P}}[f]-E_P[f]\right\} = E_{P^\eta}[f]-E_P[f] = \Xi(P^\eta||P;f). \tag{2.12}$$

The measure P^η has the form

$$dP^\eta = e^{cf-\Lambda_P^f(c)}dP, \tag{2.13}$$

where $c = c(\eta)$ is the unique nonnegative solution of $R(P^\eta||P) = \eta$.

Proof. Items 1 and 2 are proved in [23]; see also [43] for item 2. Various versions of the proof of item 3 can be found in [15] and [23]. See Proposition 3 in [24] for a more detailed statement of the result; see also similar results in [10, 9]. ∎

The tightness property in Proposition 2.5 is very attractive and ultimately relies on the presence of the cumulant generating function $\Lambda_P^{\widehat{f}}(c)$, which encodes the entire law of f. However, this generally makes the bound very difficult or impossible to compute explicitly; we will need to weaken (2.9) to obtain more usable bounds. Functional inequalities are one tool we will employ (see section 4). Another ingredient, which we discuss next, will be explicit bounds on the optimization problem in the definition of $\Xi^\pm(\Lambda, \eta)$. Such an approach was put forward in [33], where various concentration inequalities such as Hoeffding, sub-Gaussian, and Bennett bounds are discussed. For this paper we will almost exclusively use the following Bernstein-type bound.

Lemma 2.6. *Suppose there exist $\sigma > 0$, $M^\pm \geq 0$ such that*

$$\Lambda(\pm c) \leq \frac{\sigma^2 c^2}{2(1 - cM^\pm)} \tag{2.14}$$

for all $0 < c < 1/M^\pm$. Then for all $\eta \geq 0$ we have

$$\Xi^\pm(\Lambda, \eta) \leq \sqrt{2\sigma^2\eta} + M^\pm\eta. \tag{2.15}$$

Note that $M^\pm = 0$ covers the case of a (one-sided) sub-Gaussian concentration bound.

Proof. Bound Λ using (2.14), and solve the resulting optimization problem on $0 < c < 1/M^\pm$. ∎

From the point of view of concentration inequalities, the bound (2.14) is not very tight; indeed, it holds for the cumulant generating function $\Lambda_P^{\widehat{f}}$ of any random variable $f \in \mathcal{E}(P)$, but explicit constants may be hard to come by. In the context of Markov processes, however, it has proved to be extremely useful; see [56, 45, 12, 34] and in particular [29].

Second, we will need a linearization bound, generalizing (2.11).

Lemma 2.7. *Let $\Lambda : \mathbb{R} \to [0, \infty]$ be C^2 on a neighborhood of 0, $\Lambda(0) = \Lambda'(0) = 0$, and $\Lambda''(0) > 0$. Then*

$$\inf_{c>0} \left\{ \frac{\Lambda(\pm c) + \eta}{c} \right\} \leq \sqrt{2\Lambda''(0)\eta} + o(\sqrt{\eta}) \tag{2.16}$$

as $\rho \searrow 0$. If Λ'' is Lipschitz at 0, then the error bound improves to $O(\eta)$.

Proof. The bound follows from Taylor expansion of $\Lambda(c)$; see the proof of Theorem 2.8 in [23]. ∎

2.3. UQ for Markov processes. One of the main advantages of the Gibbs information inequality, (2.7), over classical information inequalities (such as the Kullback–Leibler–Cziszàr inequality) is how it scales with time when applied to the distributions of processes on path space. See [41] for a detailed discussion of this issue. This strength will become apparent as we proceed.

The following assumption details the setting in which we will work for the remainder of this paper.

Assumption 2.8. *Let $\mathcal{X}$ be a Polish space, and suppose we have a time homogeneous, $\mathcal{X}$-valued, càdlàg Markov family $(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, P^x)$, $x \in \mathcal{X}$, with transition probability kernel p_t (see the statement of Theorem C.1 in Appendix C for the precise definition of a Markov family that we use).*

Also assume we have a second probability kernel $\widetilde{P}^x$, $x \in \mathcal{X}$, on $(\Omega, \mathcal{F})$ with $(X_0)_*\widetilde{P}^x = \delta_x$ for each $x \in \mathcal{X}$.

Remark 2.9. We are *not* assuming $(X_t, \widetilde{P}^x)$ are Markov processes.

One of the families, P^x or $\widetilde{P}^x$, is thought of as the base model, and the other as some alternative (or approximate) model, but which is which can vary with the application. From a mathematical perspective, the primary factors distinguishing P^x and $\widetilde{P}^x$ are as follows:

1. Our methods require information on the spectrum of the generator of the semigroup associated with p_t.
2. (X_t, P^x) must be Markov, but $(X_t, \widetilde{P}^x)$ can be non-Markovian.

P^x and $\widetilde{P}^x$ should be chosen with these points in mind; in the remainder of this paper, we will refer to the former as the base model and the latter as the alternative model.

Definition 2.10. *Given initial distributions μ and $\widetilde{\mu}$ on $\mathcal{X}$, we also define the probability measures*

$$P^\mu(\cdot) = \int P^x(\cdot)\mu(dx), \quad \widetilde{P}^{\widetilde{\mu}}(\cdot) = \int \widetilde{P}^x(\cdot)\widetilde{\mu}(dx). \tag{2.17}$$

Note that Assumption 2.8 implies that X_t is a Markov process for the space $(\Omega, \mathcal{F}_t, P^\mu)$ with initial distribution μ and time-homogeneous transition probabilities p_t.

We will also need the finite-time restrictions, which can be thought of as the distributions on path space up to some $T > 0$,

$$P_T^x \equiv P^x|_{\mathcal{F}_T}, \ \widetilde{P}_T^x \equiv \widetilde{P}^x|_{\mathcal{F}_T}, \tag{2.18}$$

and similarly for P_T^μ and $\widetilde{P}_T^{\widetilde{\mu}}$. Finally, we let E^μ denote the expected value with respect to P^μ, and similarly for $\widetilde{E}^{\widetilde{\mu}}$.

Now fix a bounded measurable $f : \mathcal{X} \to \mathbb{R}$ (the boundedness assumption will be relaxed later) and an invariant measure μ^* for p_t. As mentioned in the introduction, there are many classical techniques for studying convergence of the ergodic averages of f under P^μ to the average in the invariant measure, $\mu^*[f]$. Therefore, in this paper we consider the much less studied problem of bounding the bias when the finite-time averages are computed by sampling from the alternative distribution; see (1.2).

2.4. UQ bounds via the Feynman–Kac semigroup. Due to our interest in the problem (1.2), we start the P-process in the invariant distribution μ^*, while the $\widetilde{P}$-process is started in an arbitrary distribution $\widetilde{\mu}$.

Given a bounded measurable function f on $\mathcal{X}$ and $T > 0$, define the bounded and $\mathcal{F}_T$-measurable function

$$f_T = \int_0^T f(X_t)dt. \tag{2.19}$$

Applying the Gibbs information inequality, (2.7), to f_T, $\widetilde{P}_T^{\widetilde{\mu}}$, $P_T^{\mu^*}$ and dividing by T yields the following theorem.

Theorem 2.11.

$$\pm\left(\widetilde{E}^{\widetilde{\mu}}\left[\frac{1}{T}\int_0^T f(X_t)dt\right]-\mu^*[f]\right)\leq \Xi^{\pm}\left(\frac{1}{T}\Lambda_{P_T^{\mu^*}}^{\widehat{f}_T},\frac{1}{T}R(\widetilde{P}_T^{\widetilde{\mu}}||P_T^{\mu^*})\right), \tag{2.20}$$

where

$$\mu^*[f]\equiv\int f d\mu^*,\quad \widehat{f}_T\equiv\int_0^T f(X_t)-\mu^*[f]dt. \tag{2.21}$$

Remark 2.12. Recall the definition

$$\Xi^{\pm}(\Lambda,\eta)=\inf_{c>0}\left\{\frac{\Lambda(\pm c)+\eta}{c}\right\}. \tag{2.22}$$

All of the UQ bounds we obtain will be of the form

$$\pm\left(\widetilde{E}^{\widetilde{\mu}}\left[\frac{1}{T}\int_0^T f(X_t)dt\right]-\mu^*[f]\right)\leq \Xi^{\pm}(\Lambda,\eta) \tag{2.23}$$

for some $\Lambda:\mathbb{R}\to[0,\infty]$ and $\eta>0$; we will refer back to these equations often.

To produce a more explicit bound from (2.20), one needs to bound the cumulant generating function as well as the relative entropy. The latter will be addressed in section 6. As for the former, observe that the cumulant generating function can be written

$$\Lambda_{P_T^{\mu^*}}^{\widehat{f}_T}(\pm c)=\log\left(\int E^x\left[\exp\left(\pm c\int_0^T f(X_t)-\mu^*[f]dt\right)\right]\mu^*(dx)\right). \tag{2.24}$$

Equation (2.24) is related to the Feynman–Kac semigroup on $L^2(\mu^*)$ with potential V:

$$\mathcal{P}_t^V[g](x)=E^x\left[g(X_t)\exp\left(\int_0^t V(X_s)ds\right)\right]. \tag{2.25}$$

More specifically,

$$\Lambda_{P_T^{\mu^*}}^{\widehat{f}_T}(\pm c)\leq\log\left(\|\mathcal{P}_T^{V_{\pm c}}[1]\|_{L^2(\mu^*)}\right), \tag{2.26}$$

$$V_{\pm c}(x)\equiv\pm\, c\,(f(x)-\mu^*[f]), \tag{2.27}$$

and so we obtain the following lemma.

Lemma 2.13. *Under Assumption 2.8, for any bounded measurable $f : \mathcal{X} \to \mathbb{R}$, (2.23) holds with*

$$\Lambda(\pm c) = \frac{1}{T}\log\left(\|\mathcal{P}_T^{V_{\pm c}}[1]\|_{L^2(\mu^*)}\right), \quad \eta = \frac{1}{T}R(\widetilde{P}_T^{\widetilde{\mu}}\|P_T^{\mu^*}). \tag{2.28}$$

In the following two sections, we discuss how functional inequalities can be used to obtain more explicit bounds on the norm of the Feynman–Kac semigroup.

3. Bounding the Feynman–Kac semigroup. The Lumer–Phillips theorem (a variant of the Hille–Yosida theorem) is our tool of choice for bounding the Feynman–Kac semigroup; see Chapter IX, page 250 in [57] or Corollary 3.20 in Chapter II of [25]. This is the same strategy used in [56, 12, 29] to obtain concentration inequalities.

First we state some of the basic properties of the Feynman–Kac semigroup, adapted from [56, 12].

Proposition 3.1. *Let $V : \mathcal{X} \to \mathbb{R}$ be bounded and measurable and μ^* be an invariant probability measure for p_t. The operators $\mathcal{P}_t^V$, $t \geq 0$, defined in (2.25), are bounded linear operators on $L^2(\mu^*)$ that form a strongly continuous semigroup.*

If $(A, D(A))$ denotes the generator of $\mathcal{P}_t \equiv \mathcal{P}_t^0$ on $L^2(\mu^*)$, then the generator of $\mathcal{P}_t^V$ on $L^2(\mu^*)$ is $(A + V, D(A))$.

Remark 3.2. $D(A)$ consists of complex-valued functions. We will use $D(A, \mathbb{R})$ to denote the real-valued functions in the domain of A.

To bound the norm of the Feynman–Kac semigroup, we use the following Hilbert space version of the Lumer–Phillips theorem (again, see [57, 25] as well as Theorem II.3.23 in [25] for a proof that (3.1) implies $A - \alpha$ is dissipative).

Proposition 3.3. *Let H be a Hilbert space and $Q(t)$ be a strongly continuous semigroup on H with generator $(A, D(A))$. Suppose that there is an $\alpha \in \mathbb{R}$ such that*

$$\mathrm{Re}(\langle Ax, x\rangle) \leq \alpha \tag{3.1}$$

for all $x \in D(A)$ with $\|x\| = 1$. Then $\|Q(t)\| \leq e^{\alpha t}$ for all $t \geq 0$.

Propositions 3.1 and 3.3 together yield a bound on the Feynman–Kac semigroup, in terms of the generator; this result, and generalizations, were proven in [56] (see Case I in the proof of Theorem 1).

Proposition 3.4. *Let $V : \mathcal{X} \to \mathbb{R}$ be bounded and measurable, and for $t \geq 0$ consider the Feynman–Kac semigroup $\mathcal{P}_t^V : L^2(\mu^*) \to L^2(\mu^*)$ with generator $(A + V, D(A))$.*

Define

$$\kappa(V) = \sup\left\{\mathrm{Re}(\langle (A+V)[g], g\rangle) : g \in D(A), \|g\|_{L^2(\mu^*)} = 1\right\} \tag{3.2}$$

$$= \sup\left\{\langle A[g], g\rangle + \int V|g|^2 d\mu^* : g \in D(A, \mathbb{R}),\ \|g\|_{L^2(\mu^*)} = 1\right\}, \tag{3.3}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product on $L^2(\mu^*)$.

Then the operator norm satisfies the bound

$$\|\mathcal{P}_t^V\| \le e^{t\kappa(V)} \tag{3.4}$$

for all $t \ge 0$.

Combining (3.4) with (2.26) and (2.20), we obtain UQ bounds that are expressed in terms of the generator of the dynamics of the baseline process and the relative entropy of the alternative process with respect to the base.

Corollary 3.5. *Under Assumption 2.8, for any bounded measurable $f : \mathcal{X} \to \mathbb{R}$, the UQ bound (2.23) holds with*

$$\Lambda(\pm c) = \kappa(V_{\pm c}), \quad \eta = \frac{1}{T} R(\widetilde{P}_T^{\widetilde{\mu}} || P_T^{\mu^*}). \tag{3.5}$$

From (2.23) we see that functional inequalities, by which we mean bounds on the generator A that lead to bounds on $\kappa(V_{\pm c})$, can be used to produce UQ bounds. Also, note that the only remaining T-dependence is in the relative entropy term, $R(\widetilde{P}_T^{\widetilde{\mu}}||P_T^{\mu^*})/T$. This will often have a finite limit (the relative entropy rate) as $T \to \infty$; for examples, see section 6.2 as well as [30], the supplementary materials to [23], and Appendix 1 of [42]. Hence Corollary 3.5 shows that one can expect UQ bounds that are well behaved as $T \to \infty$.

Remark 3.6. Proposition 3.4 is stated for bounded V, but it can be extended to certain unbounded V under the additional assumption that the symmetrized Dirichlet form is closable; see Theorem 1 in [56]. However, as noted in Corollary 3 in this same reference (and outlined in Proposition 4.12 below), that assumption can be avoided in the presence of functional inequalities by working with bounded V and then taking limits; this is the strategy we employ here.

4. UQ bounds from functional inequalities.

In this section, we explore the consequences of several important classes of functional inequalities: Poincaré, log-Sobolev, and Liapunov functions. Discussion of F-Sobolev inequalities, a generalization of the classical log-Sobolev case, can be found in Appendix B.

4.1. Poincaré inequality.

First we consider the case where the generator satisfies a Poincaré inequality with constant $\alpha > 0$, meaning

$$\mathrm{Var}_{\mu^*}[g] \le -\alpha \langle A[g], g\rangle \tag{4.1}$$

for all $g \in D(A, \mathbb{R})$. This can equivalently be written

$$\mathrm{Re}(\langle A[g], g\rangle) \le -\alpha^{-1} \|P^\perp g\|^2 \tag{4.2}$$

for all $g \in D(A)$, where $P^\perp$ is the orthogonal projector onto $1^\perp$.

In the presence of a Poincaré inequality, Proposition 3.4 is most useful when combined with the following perturbation result. A version of this result is contained in [56], but we present it here in a slightly more general form. The proof is given in Appendix A.

Lemma 4.1. *Let H be a Hilbert space, $A : D(A) \subset H \to H$ be a linear operator, and $B : H \to H$ be a bounded self-adjoint operator. Suppose there exist $D > 0$ and $x_0 \in H$ with $\|x_0\| = 1$ such that*

$$\langle Bx_0, x_0\rangle = 0 \quad \text{and} \quad \mathrm{Re}(\langle Ax, x\rangle) \leq -D\|P^\perp x\|^2 \tag{4.3}$$

for all $x \in D(A)$, where $P^\perp$ is the orthogonal projector onto $x_0^\perp$.

Define

$$B^+ \equiv \max\left\{\sup_{\|y\|=1} \langle By, y\rangle, 0\right\}. \tag{4.4}$$

Then for any $0 \leq c < D/B^+$ we have

$$\sup_{x\in D(A),\|x\|=1} \mathrm{Re}(\langle (A + cB)x, x\rangle) \leq \frac{c^2\|Bx_0\|^2}{D - cB^+}. \tag{4.5}$$

Remark 4.2. The above lemma applies to a general Hilbert space. In this paper, we will apply it to $H = L^2(\mu^*)$ (with the associated L^2-inner product), and $x_0 = 1$ (constant function), in which case $P^\perp[f] = f - \mu^*[f]$.

The multiplication operator by $V_{\pm 1}$ is a bounded self-adjoint operator, and $\langle V_{\pm 1}1, 1\rangle = 0$. Therefore, Lemma 4.1 implies the following lemma.

Lemma 4.3. *For all $0 \leq c < 1/(\alpha\|(f - \mu^*[f])^\pm\|_\infty)$ we have*

$$\kappa(V_{\pm c}) \leq \frac{\alpha\,\mathrm{Var}_{\mu^*}[f]c^2}{1 - \alpha\|(f - \mu^*[f])^\pm\|_\infty c}. \tag{4.6}$$

From this, combined with Corollary 3.5, we obtain the following UQ bound.

Theorem 4.4. *Under Assumption 2.8, if A satisfies the Poincaré inequality, (4.1), then for any bounded measurable $f : \mathcal{X} \to \mathbb{R}$ the bounds (2.23) and (2.15) hold with*

$$M^\pm = \alpha\|(f - \mu^*[f])^\pm\|_\infty, \quad \sigma^2 = 2\alpha\,\mathrm{Var}_{\mu^*}[f], \quad \eta = \frac{1}{T}R(\widetilde{P}_T^{\widetilde{\mu}}\|P_T^{\mu^*}). \tag{4.7}$$

4.2. Poincaré inequality for reversible processes. When the combination of μ^* and p_t is reversible, i.e., the generator A is self-adjoint on $L^2(\mu^*)$, and if a Poincaré inequality, (4.1), also holds with constant $\alpha > 0$, then one can obtain a UQ bound in terms of the asymptotic variance of f instead of the variance of f under μ^*.

First, define the Poisson operator

$$L : f \to \int_0^\infty \mathcal{P}_t[f]dt, \tag{4.8}$$

a bounded linear operator on $L_0^2(\mu^*) \equiv \{f \in L^2(\mu^*) : \mu^*[f] = 0\}$ with norm bound $\|L\| \leq \alpha$. The asymptotic variance of $f \in L^2(\mu^*, \mathbb{R})$ is defined by

$$\sigma^2(f) \equiv \langle 2L[f - \mu^*[f]], f - \mu^*[f]\rangle = 2\int_0^\infty \left(\int \mathcal{P}_t[f]f d\mu^* - (\mu^*[f])^2\right) dt. \tag{4.9}$$

Note that $0 \leq \sigma^2(f) \leq 2\alpha \operatorname{Var}_{\mu^*}[f]$.

Using these objects, one can obtain the following Bernstein-type bound. A simple proof appears below Remark 2.3 in [29]; we outline the essential ideas below. See [45] and [34] for similar earlier results.

Lemma 4.5. *For all $0 < c < 1/(\alpha\|(f-\mu^*[f])^\pm\|_\infty)$ we have*

$$\kappa(V_{\pm c}) \leq \frac{\sigma^2(f)c^2}{2(1-\alpha\|(f-\mu^*[f])^\pm\|_\infty c)}. \tag{4.10}$$

Proof. The cases where $\sigma^2(f) = 0$ or one of $\|(f-\mu^*[f])^\pm\|_\infty = 0$ are trivial, so suppose not. Using the self-adjoint functional calculus, one can see that L inverts A on $D(A) \cap L^2_0(\mu^*)$ and

$$\left|\int f g d\mu^*\right| \leq \left(\int -A[g] g d\mu^*\right)^{1/2} \left(\int -L[f] f d\mu^*\right)^{1/2} \tag{4.11}$$

for all real-valued $f \in L^2_0(\mu^*)$, $g \in D(A, \mathbb{R})$.

Hence, for any $g \in D(A, \mathbb{R})$ with $\|g\|_{L^2(\mu^*)} = 1$ and any bounded, measurable V (not necessarily related to f at this point),

$$\begin{aligned}\int V g^2 d\mu^* &= \int V(g-\mu^*[g])^2 d\mu^* + 2\mu^*[g]\int (V-\mu^*[V]) g d\mu^* + \mu^*[V]\mu^*[g]^2\\ &\leq \|V^+\|_\infty \operatorname{Var}_{\mu^*}[g] + \sqrt{2\sigma^2(V)}\sqrt{\langle -A[g], g\rangle} + \mu^*[V].\end{aligned} \tag{4.12}$$

Using the Poincaré inequality and solving for $\langle -A[g], g\rangle$ gives

$$\begin{aligned}&\langle -A[g], g\rangle \geq h\left(\int (V-\mu^*[V]) g^2 d\mu^*\right),\\ &h(r) \equiv 1_{r\geq 0}\frac{\sigma^2(V)}{2(M^\pm)^2}\left(\left(1+\frac{2M^+}{\sigma^2(V)}r\right)^{1/2} - 1\right)^2, \quad M^+ \equiv \alpha\|V^+\|_\infty.\end{aligned} \tag{4.13}$$

Letting $V = V_{\pm 1} = \pm(f-\mu^*[f])$ in (4.13) and using the result to bound κ, (3.3), results in

$$\kappa(V_{\pm c}) \leq \sup_{r\in\mathbb{R}}\{cr - h(r)\}. \tag{4.14}$$

Inequality (4.10) then follows from solving the optimization problem. ∎

As with Theorem 4.4, the Bernstein-type bound, (4.10), implies a UQ bound.

Theorem 4.6. *Under Assumption 2.8, if the generator satisfies the Poincaré inequality (4.1) and is self-adjoint on $L^2(\mu^*)$, then for any bounded measurable $f : \mathcal{X} \to \mathbb{R}$ the bounds (2.23) and (2.15) hold with*

$$M^\pm = \alpha\|(f-\mu^*[f])^\pm\|_\infty, \quad \sigma^2 = \sigma^2(f), \quad \eta = \frac{1}{T} R(\widetilde{P}_T^{\widetilde{\mu}} \| P_T^{\mu^*}). \tag{4.15}$$

Other variations can be derived using a Liapunov function. First we need a couple of definitions, taken from section 4 of [29]. Also, see this reference for further Liapunov function results that could likely be adapted to produce UQ bounds.

Definition 4.7. *A measurable function $G : \mathcal{X} \to \mathbb{R}$ is in the μ^*-extended domain of the generator, $D_{e,\mu^*}(A)$, if there is some measurable $g : \mathcal{X} \to \mathbb{R}$ such that $\int_0^t |g|(X_s)ds < \infty$ P^{μ^*}-a.s. and one P^{μ^*}-version of*

$$M_t(G) \equiv G(X_t) - G(X_0) - \int_0^t g(X_s)ds \tag{4.16}$$

is a local P^{μ^*}-martingale.

$U \in D_{e,\mu^*}(A)$ is called a Liapunov function if $U \geq 1$ and there exist a measurable $\phi : \mathcal{X} \to (0, \infty)$ and $b > 0$ such that

$$-\frac{A[U]}{U} \geq \phi - b \;\; \mu^*\text{-a.s.} \tag{4.17}$$

As shown in [29], given a Liapunov function, one can derive a bound on $\kappa(V_{\pm c})$; our method then produces a corresponding UQ bound.

Theorem 4.8. *In addition to Assumption 2.8, assume the generator, A, is self-adjoint on $L^2(\mu^*)$ and satisfies the Poincaré inequality (4.1) and that we have a Liapunov function U with $-A[U]/U \geq \phi - b$.*

Given an observable $f \in L^2(\mu^*, \mathbb{R})$ with $\|(f - \mu^*[f])^{\pm}/\phi\|_\infty < \infty$, we have the UQ bounds (2.23) and (2.15), where

$$M^{\pm} = (1 + \alpha b)\|(f - \mu^*[f])^{\pm}/\phi\|_\infty, \quad \sigma^2 = \sigma^2(f), \quad \eta = \frac{1}{T} R(\widetilde{P}_T^{\widetilde{\mu}} || P_T^{\mu^*}). \tag{4.18}$$

Proof. First let V be a bounded measurable function. This part of the proof proceeds similarly to that of Lemma 4.5, but rather than taking the supremum of V^+ in (4.12), one instead uses (4.17) to compute the following bound, where $g \in D(A, \mathbb{R})$ with $\|g\|_{L^2(\mu^*)} = 1$:

$$\begin{aligned} \int V g^2 d\mu^* \leq \mu^*[V] + \sqrt{2\sigma^2(V)}\sqrt{\langle -A[g], g\rangle} \\ + \|V^+/\phi\|_\infty \int \left(-\frac{A[U]}{U} + b\right)(g - \mu^*[g])^2 d\mu^*. \end{aligned} \tag{4.19}$$

Next, use the bound found in Lemma 5.6 in [34],

$$\int -\frac{A[U]}{U}(g - \mu^*[g])^2 d\mu^* \leq \langle -A[g], g\rangle, \tag{4.20}$$

and proceed as in Lemma 4.5 to obtain

$$\kappa(\pm cV) \leq \pm c\mu^*[V] + \frac{\sigma^2(V)c^2}{2(1 - (1 + \alpha b)\|V^{\pm}/\phi\|_\infty c)} \tag{4.21}$$

for all $0 < c < 1/((1+\alpha b)\|V^{\pm}/\phi\|_\infty)$. If f is bounded, then applying this to $V = f - \mu^*[f]$ and using Corollary 3.5 and Lemma 2.6 gives the claimed UQ bound.

For general $f \in L^2(\mu^*, \mathbb{R})$ with $\|(f - \mu^*[f])^{\pm}/\phi\|_\infty < \infty$, we employ a similar method to Corollary 3 in [56]: Define $V = f - \mu^*[f]$ and $V^n = V1_{|V|<n}$ (not to be confused with the nth power of V). Applying the above result to V^n and then using Fatou's lemma and L^2-continuity of the asymptotic variance gives

$$
\begin{aligned}
\frac{1}{T}\Lambda^{\widehat{f}_T}_{P^{\mu^*}_T}(\pm c) \leq & \frac{1}{T}\log\left(\|\mathcal{P}^{V_{\pm c}}_T[1]\|\right) \leq \liminf_{n\to\infty}\frac{1}{T}\log\left(\|\mathcal{P}^{\pm cV^n}_T[1]\|\right) \\
\leq & \liminf_{n\to\infty}\left(\pm c\mu^*[V^n] + \frac{\sigma^2(V^n)c^2}{2(1-(1+\alpha b)\|(f-\mu^*[f])^{\pm}/\phi\|_\infty c)}\right) \\
= & \frac{\sigma^2(f)c^2}{2(1-(1+\alpha b)\|(f-\mu^*[f])^{\pm}/\phi\|_\infty c)}.
\end{aligned} \tag{4.22}
$$

Having extended the bound on the cumulant generating function to such f, the claimed UQ bound follows from Proposition 2.2. ■

4.3. Poincaré inequality examples.

The study of Poincaré inequalities has a long history which we do not attempt to recount here. For a detailed discussion, see [55], which covers Poincaré inequalities for both continuous-time Markov chains and diffusions. Criteria for diffusions can also be found, for example, in [2, 3].

The following example illustrates that the Bernstein-type bounds used in this paper can be sharp for Markov processes.

4.3.1. A simple Liapunov example: The $M/M/\infty$ queue.

Following [29], let us consider the (simple) example of an $M/M/\infty$ queuing system which has infinitely many servers, each with a service rate ρ and an arrival rate λ. The state space is $\mathbb{N}$, and the generator is given by

$$
A[f](n) = \lambda f(n+1) - (\lambda + \rho n)f(n) + \rho n f(n-1). \tag{4.23}
$$

The invariant measure μ^* is a Poisson distribution with parameter λ/ρ. An explicit computation shows (see, e.g., [14]) that $\mathrm{Var}_{\mu^*}[\mathcal{P}_t f] \leq e^{-2\rho t}\,\mathrm{Var}_{\mu^*}[f]$, and thus the Poincaré constant is $1/\rho$.

To construct a Liapunov function take $U(n) = \kappa^n$ with $\kappa > 1$; we then have

$$
-\frac{A[U]}{U}(n) = \rho n(1-\kappa^{-1}) - \lambda(\kappa - 1)\,, \tag{4.24}
$$

and we can apply Theorem 4.8 to any function f with $|f| \leq C(n+\delta)$ for some $\delta > 0$.

It is instructive to consider further the case of the mean number of customers in the queue, i.e., $f = n$ and $\widehat{f} = f - \mu^*[f] = n - \lambda/\rho$. From (4.23) we obtain

$$
(A + \rho(1-\kappa^{-1})\widehat{f})[U](n) = \lambda\frac{(\kappa-1)^2}{\kappa}U(n), \tag{4.25}
$$

and thus U is an eigenvector for $A + \rho(1-\kappa^{-1})\widehat{f}$ with eigenvalue $\lambda\frac{(\kappa-1)^2}{\kappa}$. By the Perron–Frobenius theorem and Rayleigh's principle we obtain that

$$
\Lambda(c) \equiv \lim_{T\to\infty} T^{-1}\Lambda^{\widehat{f}_T}_{P^{\mu^*}_T}(c) \tag{4.26}
$$

is the maximal eigenvalue of $A + c\widehat{f}$, and thus $\Lambda\left(\rho(1-\kappa^{-1})\right) = \lambda\frac{(\kappa-1)^2}{\kappa}$ or equivalently $\Lambda(c) = \frac{\lambda c^2}{\rho^2(1-c\rho^{-1})}$. Since $A\widehat{f}(n) = \lambda - \rho n$, we can solve the Poisson equation $(-A)^{-1}\widehat{f} = \widehat{f}/\rho$, and thus the asymptotic variance is $\sigma^2(f) = 2\langle(-A)^{-1}\widehat{f}, \widehat{f}\rangle = 2\rho^{-1}\operatorname{Var}_{\mu^*}[f] = 2\lambda/\rho^2$. As a consequence we have

$$\Lambda(c) = \frac{\sigma^2(f)c^2}{2(1-c\rho^{-1})}, \tag{4.27}$$

which shows that Bernstein bounds can be sharp in the context of Markov processes, contrary to the i.i.d. setting.

4.3.2. Poincaré inequality from exponential convergence. It is well known that, when the generator, A, is self-adjoint, a Poincaré inequality is equivalent to exponential convergence in the $L^2(\mu^*)$-norm. Here, we discuss a method for deriving a Poincaré inequality from exponential convergence in alternative norms.

First, note that one only needs exponential L^2-convergence on a subset with dense span to conclude a Poincaré inequality (see Lemma 1.2 in [13]).

Lemma 4.9. *Suppose $(A, D(A))$ is self-adjoint, $F \subset L^2(\mu^*)$ has dense span, and there exists $\alpha > 0$ such that the following holds: For every $f \in F$ there exists $C_f \geq 0$ such that*

$$\|\mathcal{P}_t[f] - \mu^*[f]\|_2 \leq C_f e^{-t/\alpha} \quad \text{for all } t \geq 0. \tag{4.28}$$

Then a Poincaré inequality, (4.1), holds with constant α.

The following result shows how to obtain a Poincaré inequality (with an explicit constant) from exponential convergence in a pair of weighted norms.

Theorem 4.10. *Suppose $(A, D(A))$ is self-adjoint, and $W : \mathcal{X} \to [1,\infty)$ is measurable. Define the following norms on measurable functions $\phi : \mathcal{X} \to \mathbb{R}$ and signed measures π on $\mathcal{X}$:*

$$|\phi|_W = \sup_{x\in\mathcal{X}} \frac{|\phi(x)|}{W(x)}, \quad |\pi|_W = \int W d|\pi|. \tag{4.29}$$

Suppose we have $\lambda \geq 0, \rho \geq 0$ with at least one nonzero and that for every bounded measurable $h : \mathcal{X} \to [0,\infty)$ with $\int h d\mu = 1$ there exist $C_h, D_h \in [0,\infty)$ such that for all $t \geq 0$

$$|\mathcal{P}_t[h] - 1|_W \leq D_h e^{-\rho t}, \tag{4.30}$$

and the measure $d\nu = h d\mu^*$ satisfies

$$|\mathcal{P}_t^\dagger[\nu] - \mu^*|_W \leq C_h e^{-\lambda t}, \tag{4.31}$$

where $\mathcal{P}_t^\dagger$ denotes the action of the semigroup p_t on measures.

Then A satisfies the Poincaré inequality

$$\operatorname{Var}_{\mu^*}[g] \leq -\frac{2}{\lambda+\rho}\langle A[g], g\rangle \quad \text{for all } g \in D(A,\mathbb{R}). \tag{4.32}$$

Proof. The proof is similar to that of Theorem 2.1 in [3]. The key is to take h as above, let $d\nu = hd\mu^*$, use symmetry of $\mathcal{P}_t$ to compute

$$\|\mathcal{P}_t[h]-1\|_2^2 = \int \frac{|\mathcal{P}_t[h]-1|}{W} W|\mathcal{P}_t[h]-1|d\mu^* \leq |\mathcal{P}_t[h]-1|_W |\mathcal{P}_t^\dagger[\nu]-\mu^*|_W, \tag{4.33}$$

and then apply Lemma 4.9. ■

Exponential convergence in norms of the form $|\cdot|_W$ can be obtained from the existence of an appropriate Liapunov function (see [35, 36]), making Theorem 4.10 a practical method for obtaining Poincaré inequalities.

Remark 4.11. The proof of Lemma 4.9 can be generalized to only require (4.28) to hold along a sequence t_n^f converging to ∞. Hence, Theorem 4.10 can also be generalized to only require (4.30) and (4.31) along a common sequence $t_n^h \to \infty$.

4.4. log-Sobolev inequalities. Next consider the log-Sobolev inequality with constant $\beta > 0$:

$$\int g^2 \log(g^2) d\mu^* \leq -\beta \int A[g] g d\mu^* \tag{4.34}$$

for all $g \in D(A,\mathbb{R})$ with $\|g\|_{L^2(\mu^*)} = 1$.

We will employ the following generalization of the Feynman–Kac semigroup for (possibly) unbounded potentials. The subsequent proposition was shown in Corollary 4 in [56]. For completeness purposes, we outline the proof.

Proposition 4.12. *Let A be the generator of $\mathcal{P}_t$ and μ^* be an invariant measure for the adjoint semigroup, $\beta > 0$, and assume the* log*-Sobolev inequality,* (4.34)*, holds for μ^* with constant β.*

Finally, suppose that $V \in L^1(\mu^*)$ with $\int e^{\beta V} d\mu^* < \infty$. Then $\mathcal{P}_t^V : L^2(\mu^*) \to L^2(\mu^*)$, defined by

$$\mathcal{P}_t^V[g](x) = E^x\left[g(X_t)\exp\left(\int_0^t V(X_s)ds\right)\right], \tag{4.35}$$

are well-defined linear operators, and the operator norm satisfies the bound

$$\|\mathcal{P}_t^V\| \leq \left(\int e^{\beta V} d\mu^*\right)^{t/\beta}. \tag{4.36}$$

Proof. First assume V is bounded. Inequality (3.4) gives $\|\mathcal{P}_t^V\| \leq e^{t\kappa(V)}$. Applying the log-Sobolev inequality together with the Gibbs variational principle, (2.6), we obtain

$$\begin{aligned}\kappa(V) \leq& \beta^{-1} \sup\left\{-\int g^2\log(g^2)d\mu^* + \int \beta V|g|^2 d\mu^* : \|g\|_{L^2(\mu^*)} = 1\right\} \\ =& \beta^{-1} \sup_{d\nu = g^2 d\mu^* : \|g\|_2 = 1} \{E_\nu[\beta V] - R(\nu\|\mu^*)\} = \beta^{-1}\log\left(\int \exp(\beta V)\, d\mu^*\right),\end{aligned} \tag{4.37}$$

which proves the claim.

The case of unbounded V satisfying the assumptions of the theorem is obtained by letting $V^n = V1_{|V|\leq n}$ and then using Fatou's lemma, the result for bounded V, and dominated convergence to compute

$$\|\mathcal{P}_t^V\| \leq \liminf_{n\to\infty} \|\mathcal{P}_t^{V^n}\| \leq \liminf_{n\to\infty} \left(\int e^{\beta V^n} d\mu^*\right)^{t/\beta} = \left(\int e^{\beta V} d\mu^*\right)^{t/\beta}. \qquad \blacksquare$$

Using Proposition 4.12, a UQ bound of the form (2.23) can be derived that covers a class of unbounded observables.

Theorem 4.13. *In addition to Assumption 2.8, assume the* log-*Sobolev inequality,* (4.34)*, holds and we have an observable $f \in L^1(\mu^*, \mathbb{R})$ and $c_- < 0 < c_+$ such that for all $c \in (c_-, c_+)$*

$$\int \exp(\beta V_c)\, d\mu^* < \infty. \tag{4.38}$$

Then a UQ bound of the form (2.23) *holds with*

$$\Lambda(c) = \begin{cases} \frac{1}{\beta}\log\left(\int e^{\beta V_c} d\mu^*\right) & \text{if } c \in (c_-, c_+), \\ +\infty & \text{otherwise.} \end{cases} \tag{4.39}$$

In addition, the asymptotic result (2.16) *holds with*

$$\Lambda''(0) = \beta\, \mathrm{Var}_{\mu^*}[f], \quad \eta = \frac{1}{T} R(\widetilde{P}_T^{\widetilde{\mu}} || P_T^{\mu^*}). \tag{4.40}$$

Proof. The bound (4.36) implies $E^{\mu^*}[\exp(cf_T)] < \infty$ for $c \in (c_-, c_+)$; hence $f_T \in \mathcal{E}(P_T^{\mu^*})$, and the Gibbs information inequality, (2.7), applies. As in (2.26), the cumulant generating function can be bounded using the Feynman–Kac semigroup bound, (4.36). Combining this with (2.7) yields a bound of the form (2.23), with Λ as defined in (4.39). $\blacksquare$

The ideas in this section can be extended to F-Sobolev inequalities; see Appendix B.

4.4.1. Example: Diffusions. Let V be a C^2 potential, bounded below, and growing sufficiently fast at infinity. Consider the diffusion with generator $A = \Delta - \nabla V \cdot \nabla$ and invariant measure $\mu^*(dx) = e^{-V(x)}dx$. First, it is useful to note that a log-Sobolev inequality with constant β implies a Poincaré inequality with constant $\alpha = \beta/2$ [53]. In [11], the following sufficient condition for a log-Sobolev inequality was obtained.

Suppose A satisfies a Poincaré inequality with constant α (references on Poincaré inequalities can be found in section 4.3) and that

$$-C \equiv \inf_x \left\{\frac{1}{4}|\nabla V(x)|^2 - \frac{1}{2}\Delta V(x) - \pi e^2 V(x)\right\} > -\infty. \tag{4.41}$$

Then A satisfies a log-Sobolev inequality with constant

$$\beta = 3\alpha + \frac{1}{(1+\alpha|C|)\pi e^2}. \tag{4.42}$$

As a second example, if the Hessian of V is bounded below,

$$D^2V(x) \geq 2\beta^{-1}I \tag{4.43}$$

for some $\beta > 0$ (unrelated to the β in (4.42)), then a log-Sobolev inequality holds with constant β [4]. A UQ bound corresponding to the associated Poincaré inequality with constant $\alpha \equiv \beta/2$ was given in the introduction in (1.10).

5. Functional inequalities and UQ for discrete-time Markov processes.

In this section we show how the above framework can be applied to obtain UQ bounds for invariant measures of discrete-time Markov processes.

Again, let $\mathcal{X}$ be a Polish space, and suppose we have one-step transition probabilities $p(x, dy)$ and $\widetilde{p}(x, dy)$ on $\mathcal{X}$ with invariant measures μ^* and $\widetilde{\mu}^*$, respectively. Assume that $R(\widetilde{\mu}^*||\mu^*) < \infty$.

Define the bounded linear operator $\mathcal{P}$ on $L^2(\mu^*)$,

$$\mathcal{P}[f](x) \equiv \int f(y)p(x, dy), \tag{5.1}$$

and similarly for $\widetilde{\mathcal{P}}$ on $L^2(\widetilde{\mu}^*)$.

We obtain UQ bounds for expectations in μ^* and $\widetilde{\mu}^*$ by constructing continuous-time processes with these same invariant distributions. Specifically, in Appendix C (see Theorem C.1) we obtain càdlàg Markov families $(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, \{P^x\}_{x\in\mathcal{X}})$ and $(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, \{\widetilde{P}^x\}_{x\in\mathcal{X}})$, whose transition probabilities p_t and $\widetilde{p}_t$, respectively (not to be confused with p and $\widetilde{p}$), satisfy the following:

1. μ^* is invariant for p_t for all $t \geq 0$, and similarly for $\widetilde{\mu}^*$ and $\widetilde{p}_t$ (see Theorem C.2).
2. The continuous-time semigroup, $\mathcal{P}_t$, on $L^2(\mu^*)$ constructed from p_t is

$$\mathcal{P}_t = \exp(t(\mathcal{P} - I)). \tag{5.2}$$

   Specifically, $\mathcal{P}_t$ has bounded generator $A = \mathcal{P} - I$ (see Theorem C.2). Note that we will also refer to A as the generator of the discrete-time Markov process.
3. The relative entropy rate of the continuous-time process can be bounded by the relative entropy of the discrete-time process as follows:

$$R(\widetilde{P}_T^{\widetilde{\mu}^*}||P_T^{\mu^*}) \leq R(\widetilde{\mu}^*||\mu^*) + T\int R(\widetilde{p}(x,\cdot)||p(x,\cdot))\widetilde{\mu}^*(dx) \tag{5.3}$$

   for all $T > 0$ (see Theorem C.4 and Corollary C.5).

Remark 5.1. While the above construction, and the computation of the relative entropy, is standard for countable state spaces (see the discussion in section 6.1), for our purposes it is necessary to work with general state spaces; to the best of our knowledge, the relative entropy bound (5.3) is new in this case.

General state spaces are of interest, for example, when one is working with Markov chain Monte Carlo samplers, $p(x, dy)$ and $\widetilde{p}(x, dy)$, for measures, μ^* and $\widetilde{\mu}^*$, respectively, on $\mathbb{R}^n$. In this setting, to use our UQ method, one can construct the ancillary continuous-time Markov chain on $\mathbb{R}^n$, as outlined in Appendix C, and then apply the relative entropy bound (5.3).

The Markov families P^x and $\widetilde{P}^x$, obtained via the above construction, satisfy Assumption 2.8. Hence, if the generator $\mathcal{P}-I$ satisfies any of the functional inequalities covered in section 3, then the general results therein imply UQ bounds for expectations in the invariant measures μ^* and $\widetilde{\mu}^*$, with (5.3) providing a bound on the relative entropy rate.

Remark 5.2. Note that here we must take $\widetilde{\mu} = \widetilde{\mu}^*$ for the bounds to apply to the original discrete-time process; otherwise one obtains UQ bounds for ergodic averages of $f(X_t)$ under the auxiliary continuous-time Markov family.

For example, a Poincaré inequality for the generator $\mathcal{P} - I$,

$$\mathrm{Re}(\langle(\mathcal{P}-I)g, g\rangle) \leq -\alpha^{-1}\|P^{\perp}g\|^2_{L^2(\mu^*)}, \quad g \in L^2(\mu^*),\ \alpha > 0, \tag{5.4}$$

implies that for any bounded measurable $f : \mathcal{X} \to \mathbb{R}$, we have

$$\begin{aligned} &\pm\left(\widetilde{\mu}^*[f] - \mu^*[f]\right) \leq \sqrt{2\sigma^2\eta} + M^{\pm}\eta,\\ &\sigma^2 = 2\alpha\,\mathrm{Var}_{\mu^*}[f],\quad M^{\pm} = \alpha\|(f-\mu^*[f])^{\pm}\|_\infty,\\ &\eta = \int R(\widetilde{p}(x,\cdot)||p(x,\cdot))\widetilde{\mu}^*(dx). \end{aligned} \tag{5.5}$$

This follows from Theorem 4.4, after taking $T \to \infty$ (recall the assumption $R(\widetilde{\mu}^*||\mu^*) < \infty$).

We illustrate these discrete-time UQ bounds with a pair of examples.

5.1. Example: Random walk on a hypercube. Consider the symmetric random walk on the d-dimensional hypercube $\mathcal{X} = \{-1, 1\}^d$; i.e., the transition probabilities are defined by uniformly randomly selecting a coordinate, $i \in \{1, \dots, d\}$, and then independently and uniformly selecting the sign, 1 or -1, with which to update the selected component.

The uniform measure, μ^*, on $\mathcal{X}$ is invariant and the process is reversible on $(\mathcal{X}, \mu^*)$. The eigenvalues and eigenvectors of the transition matrix can be found explicitly; see Example 12.15 in [44]. In particular, the second largest eigenvalue is $\lambda_2 = 1 - 1/d$; hence we obtain the following Poincaré inequality:

$$\mathrm{Re}(\langle(\mathcal{P}-I)g, g\rangle) \leq -\frac{1}{d}\|P^{\perp}g\|^2_{L^2(\mu^*)}, \quad g \in L^2(\mu^*). \tag{5.6}$$

Assuming $R(\widetilde{\mu}^*||\mu^*) < \infty$, we then obtain the UQ bound (5.5) with $\alpha = d$.

5.2. Example: Exclusion chain. Derivation of functional inequalities for many discrete-time Markov processes can be found in [19]. Here we investigate the resulting UQ bounds for one of these examples; see section 4.6 in the above reference and also [18] for further details and proofs regarding this example.

Let (V, E) be a symmetric, connected graph with n vertices. Let $d(x)$ be the degree of a vertex $x \in V$ and $d_0 = \max_x d(x)$. Fix $r \leq n$. The r-exclusion process is a Markov chain with state space being the set of cardinality r subsets of V. Informally stated, the transition probabilities are defined as follows: Given an r-subset A (i.e., state of the chain), pick an element $x \in A$ with probability proportional to its degree. Uniformly randomly pick a vertex y out of all those connected with x. If y is not in A, then transition to the new state $(A \setminus \{x\}) \cup \{y\}$. Otherwise, the chain remains at the set A.

For each $(x,y) \in V \times V$, fix a path $\gamma_{x,y}$ from x to y in the graph, and let $|\gamma_{x,y}|$ be its length. Define

$$\Delta_0 = \max_{e_0 \in V} \left\{ \sum_{(x,y): e_0 \in \gamma_{x,y}} |\gamma_{x,y}| \right\}, \quad d_r = \max_{A \subset V: |A|=r} \left\{ \frac{1}{r} \sum_{a \in A} d(a) \right\}. \tag{5.7}$$

The generator of this Markov chain satisfies both a Poincaré inequality and a log-Sobolev inequality with respective constants being

$$\alpha = r d_r \Delta_0 / n, \quad \beta = 3 r d_r \Delta_0 \log(n)/n. \tag{5.8}$$

Then, assuming $R(\widetilde{\mu}^*||\mu^*) < \infty$, the above Poincaré inequality implies the UQ bound (5.5) with α as in (5.8), and the log-Sobolev inequality results in

$$\pm \left(\widetilde{\mu}^*[f] - \mu^*[f]\right) \le \inf_{c>0} \left\{ \frac{1}{c\beta} \log \left(\int \exp\left(\pm \beta c (f - \mu^*[f])\right) d\mu^* \right) + \frac{\eta}{c} \right\}, \tag{5.9}$$

with β and η as in (5.8) and (5.5), respectively.

6. Bounding the relative entropy rate.

For any $\eta > 0$, the results derived in the previous sections provide UQ bounds over the class of all alternative models that satisfy a relative entropy bound of the form

$$H_T(\widetilde{P}^{\widetilde{\mu}}||P^{\mu^*}) \equiv \frac{1}{T} R(\widetilde{P}_T^{\widetilde{\mu}}||P_T^{\mu^*}) \le \eta. \tag{6.1}$$

In this section, we study in more detail the dependence of H_T on T and on the models $\widetilde{P}^{\widetilde{\mu}}$ and P^{μ^*}. Specifically, we derive upper bounds on H_T in various settings that can be substituted for H_T in the general UQ bound (2.23). Here, it will make little difference whether the initial distribution for the P-process is invariant or not, so we no longer make that assumption when deriving the relative entropy bounds; μ will denote an arbitrary initial distribution.

Deriving bounds on the relative entropy is a very application-specific problem. We will cover several examples in detail: continuous-time Markov chains, semi-Markov processes, change of drift in SDEs, and numerical methods for SDEs with additive noise.

6.1. Example: Continuous-time Markov chains.

Let $\mathcal{X}$ be a countable set, P^μ, $\widetilde{P}^{\widetilde{\mu}}$ be probability measures on $(\Omega, \mathcal{F})$, and $X_t : \Omega \to \mathcal{X}$ such that P^μ (resp., $\widetilde{P}^{\widetilde{\mu}}$) makes $(\Omega, \mathcal{F}, X_t)$ a continuous-time Markov chain with transition probabilities $a(x,y)$ (resp., $\widetilde{a}(x,y)$), jump rates $\lambda(x)$ (resp., $\widetilde{\lambda}(x)$), and initial distribution μ (resp., $\widetilde{\mu}$). Let $\mathcal{F}_t$ be the natural filtration for X_t and X_n^J be the embedded jump chain with jump times J_n.

Suppose $\widetilde{\mu} \ll \mu$, λ and $\tilde{\lambda}$ are positive and bounded above, and for all $x, y \in \mathcal{X}$ we have $a(x,y) = 0$ iff $\widetilde{a}(x,y) = 0$. Then for any $T > 0$ we have $\widetilde{P}^{\widetilde{\mu}}|_{\mathcal{F}_T} \ll P^\mu|_{\mathcal{F}_T}$ and

$$\begin{aligned} &R(\widetilde{P}^{\widetilde{\mu}}|_{\mathcal{F}_T}||P^\mu|_{\mathcal{F}_T}) \\ =&R(\widetilde{\mu}||\mu) + \widetilde{E}^{\widetilde{\mu}}\left[\int_0^T \widetilde{F}(X_s)\widetilde{\lambda}(X_s) ds\right] - \widetilde{E}^{\widetilde{\mu}}\left[\int_0^T \widetilde{\lambda}(X_s) - \lambda(X_s) ds\right], \\ &\widetilde{F}(x) \equiv \sum_{z \in \mathcal{X}} \widetilde{a}(x,z) \log\left(\frac{\widetilde{\lambda}(x)\widetilde{a}(x,z)}{\lambda(x) a(x,z)}\right). \end{aligned} \tag{6.2}$$

To simplify further, if $\widetilde{\mu} = \widetilde{\mu}^*$ is an invariant measure, then

$$R(\widetilde{P}^{\widetilde{\mu}^*}|_{\mathcal{F}_T}||P^{\mu}|_{\mathcal{F}_T}) = R(\widetilde{\mu}^*||\mu) + T\left(\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{X}}\widetilde{\mu}^*(x)\widetilde{\lambda}(x)\widetilde{a}(x,z)\log\left(\frac{\widetilde{\lambda}(x)\widetilde{a}(x,z)}{\lambda(x)a(x,z)}\right) - \sum_{x\in\mathcal{X}}\widetilde{\mu}^*(x)\left(\widetilde{\lambda}(x)-\lambda(x)\right)\right). \tag{6.3}$$

See the supplementary materials to [23] and Proposition 2.6 in Appendix 1 of [42] for details regarding these results.

6.2. Example: Semi-Markov processes. As we have noted previously, our results require (X_t, P^x) to be Markov but do *not* require the alternative model $(X_t, \widetilde{P}^x)$ to be Markov. Here we discuss one such class of examples, that of a semi-Markov perturbation of a continuous-time Markov chain.

Semi-Markov processes are continuous-time jump processes with memory (i.e., with nonexponential waiting times). Such a process is defined by a jump chain, X_n^J, jump times, J_n, and waiting times (i.e., jump intervals), $\Delta_{n+1} \equiv J_{n+1} - J_n$, that satisfy

$$\begin{aligned}&\widetilde{P}^{\widetilde{\mu}}(X_{n+1}^J = y, \Delta_{n+1} \le t | X_1^J, \dots, X_{n-1}^J, X_n^J, J_1, \dots, J_n)\\ &= \widetilde{P}^{\widetilde{\mu}}(X_{n+1}^J = y, \Delta_{n+1} \le t | X_n^J) \equiv \widetilde{Q}_{X_n^J, y}(t).\end{aligned}$$

$\widetilde{Q}_{x,y}(t)$ is called the semi-Markov kernel; see, for example, [38, 48] for further details. Note that a continuous-time Markov chain with embedded jump Markov chain transition probabilities $a(x,y)$ and jump rates $\lambda(x)$ is described by the semi-Markov kernel

$$Q_{x,y}(t) = a(x,y)\int_0^t \lambda(x)e^{-\lambda(x)s}ds. \tag{6.4}$$

A semi-Markov perturbation of (6.4) with the same embedded jump Markov chain but with modified (nonexponential) waiting times is described by a kernel of the form

$$\widetilde{Q}_{x,y}(t) = a(x,y)\widetilde{H}_x(t). \tag{6.5}$$

Remark 6.1. Phase-type distributions constitute a useful semiparametric description of such alternative waiting-time distributions, going beyond the exponential case to describe systems with memory; see [27, 7] for details.

The relative entropy rate,

$$\eta \equiv \limsup_{T\to\infty}\frac{1}{T}R(\widetilde{P}^{\widetilde{\mu}}|_{\mathcal{F}_T}||P^{\widetilde{\mu}^*}|_{\mathcal{F}_T}), \tag{6.6}$$

between semi-Markov processes was obtained in [30] under the appropriate ergodicity assumptions. When the base process has the form (6.4) and the alternative process has the form (6.5), the relative entropy rate can be expressed in terms of the relative entropy of the waiting-time distributions:

$$\eta = \frac{1}{\widetilde{m}_\pi}\sum_x \pi(x)R(\widetilde{H}_x||H_x), \quad \widetilde{m}_\pi \equiv \sum_x \pi(x)\int_0^\infty (1-\widetilde{H}_x(t))dt, \tag{6.7}$$

where π is the invariant distribution for the Markov chain $a(x,y)$.

Remark 6.2. The quantity $\widetilde{m}_\pi$ is the mean sojourn time under the invariant distribution, π, and $\sum_x \pi(x)R(\widetilde{H}_x||H_x)$ can be thought of as the mean relative entropy of a single jump (comparing the alternative and base model waiting-time distributions). The formula for η, (6.7), therefore has the intuitive meaning of an information loss per unit time.

6.2.1. Semi-Markov perturbations of an $M/M/\infty$-queue. As a concrete example, we consider semi-Markov perturbations of an $M/M/\infty$-queue with service rate ρ and with an arrival rate λ. The embedded jump Markov chain is given by

$$a(x,x+1)=\lambda/(\lambda+\rho x),\quad a(x,x-1)=\rho x/(\lambda+\rho x), \tag{6.8}$$

and the waiting-times are exponentially distributed with jump rates

$$\lambda(x)=\alpha+\rho x. \tag{6.9}$$

Equation (6.8) has invariant distribution

$$\pi(x)=\frac{(\alpha+\rho x)(\alpha/\rho)^x}{2\alpha x!}e^{-\alpha/\rho}. \tag{6.10}$$

Taking $T\to\infty$ in (2.20) and using (4.26), (4.27), and (6.7), we therefore obtain the following asymptotic upper bound on the average queue length in the alternative model:

$$\begin{aligned}&\limsup_{T\to\infty}\left(\widetilde{E}^{\widetilde{\mu}}\left[\frac{1}{T}\int_0^T X_t dt\right]-\alpha/\rho\right)\\ &\le \inf_{0<c<\rho}\left\{\frac{1}{c}\frac{\alpha c^2}{\rho^2(1-c/\rho)}+\frac{1}{c}\eta\right\}=\left(2\sqrt{\eta/\alpha}+\eta/\alpha\right)\frac{\alpha}{\rho},\end{aligned} \tag{6.11}$$

where

$$\begin{aligned}\eta =&\frac{1}{\widetilde{m}_\pi}\sum_x \pi(x)R(\widetilde{H}_x||H_x),\\ \widetilde{m}_\pi =&\sum_x \pi(x)\int_0^\infty(1-\widetilde{H}_x(t))dt,\quad H_x(t)=\int_0^t\lambda(x)e^{-\lambda(x)s}ds.\end{aligned} \tag{6.12}$$

Note that the only ingredient from the alternative model that is needed in (6.11) is $\widetilde{H}_x$, and given this, the bounds are generally straightforward to evaluate.

6.3. Example: Change of drift for SDEs. Next, consider the case where P^x and $\widetilde{P}^x$ are the distributions on $C([0,\infty),\mathbb{R}^n)$ of the solution flows X_t^x and $\widetilde{X}_t^x$ of a pair of SDEs. More precisely, we have the following.

Assumption 6.3. *Assume the following:*

1. X_t^x *and* $\widetilde{X}_t^x$ *are weak solutions to the* $\mathbb{R}^n$*-valued SDEs, on filtered probability spaces satisfying the usual conditions* [40]:

$$dX_t^x=b(X_t^x)dt+\sigma(X_t^x)dW_t,\ X_0^x=x, \tag{6.13}$$

$$d\widetilde{X}_t^x=\widetilde{b}(\widetilde{X}_t^x)dt+\sigma(\widetilde{X}_t^x)d\widetilde{W}_t,\ \widetilde{X}_0^x=x, \tag{6.14}$$

where W_t and $\widetilde{W}_t$ are m-dimensional Wiener processes. We let P and $\widetilde{P}$ denote the probability measures of the respective spaces where the SDEs are defined.
Here we think of $b : \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ as the measurable drift and diffusion for the base process, and we assume the modified drift has the form $\widetilde{b} = b + \sigma\beta$ for some measurable $\beta : \mathbb{R}^n \to \mathbb{R}^m$.

2. *X_t^x and $\widetilde{X}_t^x$ are jointly continuous in (t, x).*
3. *X_t^x satisfies the following flow property: For any bounded, measurable $G : C([0,\infty), \mathbb{R}^n) \to \mathbb{R}$, we have*

$$E_P(G(X^x_{t+\cdot})|\mathcal{F}_t) = E_P\left[G\left(X^{(\cdot)}\right)\right] \circ X_t^x. \tag{6.15}$$

4. *X_t^x and β satisfy the Novikov condition*

$$E_P\left[\exp\left(\frac{1}{2}\int_0^T \|\beta(X_s^x)\|^2 ds\right)\right] < \infty \tag{6.16}$$

for all $x \in \mathbb{R}^n$, $T > 0$.

5. *For every $T > 0$, solutions to (6.14) satisfy uniqueness in law, up to time T.*

Given this, we define $P^x = (X^x)_* P$ and $\widetilde{P}^x = (\widetilde{X}^x)_*\widetilde{P}$, i.e., the distributions on path space, with the Borel sigma algebra:

$$(\Omega, \mathcal{F}, \mathcal{F}_t) = (C([0,\infty), \mathbb{R}^n), \mathcal{B}(C([0,\infty), \mathbb{R}^n)), \sigma(\pi_s, s \le t)), \tag{6.17}$$

where π_t is evaluation at time t. Finally, define $X_t \equiv \pi_t$. One can easily show that the above properties are sufficient to guarantee that Assumption 2.8 holds.

Remark 6.4. The existence of flows of solutions X_t^x and $\widetilde{X}_t^x$ that satisfy the above conditions is guaranteed, for example, if b and σ satisfy a linear growth bound

$$\|b(x)\|^2 + \|\sigma(x)\|^2 \le K^2(1 + \|x\|^2), \tag{6.18}$$

and the following local Lipschitz bound.

For each ℓ there exists K_ℓ such that

$$\|b(x) - b(y)\| + \|\sigma(x) - \sigma(y)\| \le K_\ell\|x - y\| \tag{6.19}$$

on $\|x\|, \|y\| \le \ell$, and if $\beta : \mathbb{R}^n \to \mathbb{R}^m$ is also bounded and locally Lipschitz.

Fixing $T > 0$, Girsanov's theorem allows one to bound the relative entropy, $R(\widetilde{P}_T^x||P_T^x)$, that appears in the UQ bound (2.23). See the supplementary materials to [23] for more details.

Lemma 6.5. *Under Assumption 6.3, and given initial distributions μ and $\widetilde{\mu}$ for the base and alternative models, respectively, we have*

$$H_T(\widetilde{P}^{\widetilde{\mu}}||P^\mu) \le \frac{1}{T}R(\widetilde{\mu}||\mu) + \int\left(\frac{1}{2T}\int_0^T E_{\widetilde{P}}\left[\|\beta(\widetilde{X}_s^x)\|^2\right] ds\right)\widetilde{\mu}(dx).$$

6.4. Example: Euler–Maruyama methods for SDEs with additive noise. As the final example, we consider SDEs with additive noise, approximated by a (generalized) Euler–Maruyama (EM) method.

Assumption 6.6. *Let W_t be an n-dimensional Wiener process on filtered probability spaces satisfying the usual conditions, let $b : \mathbb{R}^n \to \mathbb{R}^n$ satisfy the linear boundedness and local Lipschitz properties as described in Remark 6.4, and let X_t^x be the strong solutions to the SDEs*

$$dX_t^x = b(X_t^x)dt + dW_t, \quad X_0^x = x. \tag{6.20}$$

Recall that versions can be chosen so that X_t^x is jointly continuous in (t,x) and X_t^x satisfies the flow property (6.15).

We fix $\Delta t > 0$ and assume we are given a measurable vector field $\widetilde{b}_{\Delta t} : \mathbb{R}^n \to \mathbb{R}^n$ (the drift for the generalized EM method). We define the approximating process $\widetilde{X}_0^x = x$,

$$\widetilde{X}^x|_{(j\Delta t,(j+1)\Delta t]}(t) = \widetilde{X}^x_{j\Delta t} + \widetilde{b}_{\Delta t}(\widetilde{X}^x_{j\Delta t})(t - j\Delta t) + W_t - W_{j\Delta t} \quad \text{for } j \in \mathbb{Z}_0. \tag{6.21}$$

We emphasize that, for the purposes of employing the theory we have developed (i.e., to employ functional inequalities satisfied by the generator of (6.20)), it is necessary to extend $\widetilde{X}_t^x$ to all $t \geq 0$ and not just define it at the mesh points $j\Delta t$.

Let P denote the probability measure on the space where the SDE is defined. Similarly to the previous example, we define $P^x = (X^x)_* P$ and $\widetilde{P}^x = (\widetilde{X}^x)_* P$, probability measures on

$$(\Omega, \mathcal{F}, \mathcal{F}_t) = (C([0,\infty), \mathbb{R}^n), \mathcal{B}(C([0,\infty), \mathbb{R}^n)), \sigma(\pi_s, s \leq t)). \tag{6.22}$$

Assumption 6.6 is sufficient to guarantee that Assumption 2.8 holds. The chain rule for relative entropy (see Theorem C.3.1 in [22]) can be used to obtain

$$R(\widetilde{P}_T^{\widetilde{\mu}} || P_T^{\mu}) \leq R(\widetilde{\mu}||\mu) + \int R(\widetilde{P}_T^x || P_T^x) \widetilde{\mu}(dx). \tag{6.23}$$

Let $T = N\Delta t$ for $N \in \mathbb{Z}^+$. For the purposes of bounding the relative entropy term

$$R(\widetilde{P}_T^x || P_T^x) = R((\widetilde{X}^x|_{[0,N\Delta t]})_* P || (X^x|_{[0,N\Delta t]})_* P), \tag{6.24}$$

it will be useful to define the Polish space $\mathcal{Y} \equiv C([0,\Delta t], \mathbb{R}^n)$ and the following one-step transition probabilities for a discrete-time Markov process on $\mathcal{Y}$:

$$q(y,B) = P\left(X^{y(\Delta t)}|_{[0,\Delta t]} \in B\right), \quad \widetilde{q}(y,B) = P\left(\widetilde{X}^{y(\Delta t)}|_{[0,\Delta t]} \in B\right). \tag{6.25}$$

Letting $\otimes_1^N q$ denote the composition on $\mathcal{Y}^N$, the Markov property implies

$$\otimes_1^N q(x,\cdot) = \left(X^x|_{[0,\Delta t]}, X^x|_{\Delta t + [0,\Delta t]}, \ldots, X^x|_{(N-1)\Delta t + [0,\Delta t]}\right)_* P \tag{6.26}$$

for all $x \in \mathbb{R}^n$, and similarly for $\widetilde{q}$, $\widetilde{X}^x$.

Therefore, using the chain rule for relative entropy again, we obtain

$$R(\widetilde{P}^x_{N\Delta t} || P^x_{N\Delta t}) = \sum_{j=0}^{N-1} \int R\left(\widetilde{q}(y,\cdot)||q(y,\cdot)\right) \widetilde{q}^j(x,dy) \tag{6.27}$$

for all $x \in \mathbb{R}^n$. Hence we arrive at the following lemma.

Lemma 6.7.

$$R(\widetilde{P}^x_{N\Delta t}||P^x_{N\Delta t}) = \sum_{j=1}^N E_P\left[R\left(\widetilde{P}^{(\cdot)}_{\Delta t}||P^{(\cdot)}_{\Delta t}\right)\circ \widetilde{X}^x_{(j-1)\Delta t}\right]. \tag{6.28}$$

The one-step relative entropy can be bounded via Girsanov's theorem, similarly to Lemma 6.5; on each time interval of length Δt, the tilde process is simply the solution to an SDE with constant drift and additive noise.

Lemma 6.8. *Under Assumption* 6.6

$$\begin{aligned} H_{N\Delta t}(\widetilde{P}^{\widetilde{\mu}}||P^\mu) \leq& \frac{1}{N\Delta t}R(\widetilde{\mu}||\mu)\\ &+\frac{1}{N}\sum_{j=1}^N\int\int E_P\left[\frac{1}{2\Delta t}\int_0^{\Delta t}\|\widetilde{b}_{\Delta t}(y)-b(\widetilde{X}^x_s)\|^2 ds\right]\widetilde{p}^{\Delta t}_{j-1}(x,dy)\widetilde{\mu}(dx), \end{aligned} \tag{6.29}$$

where $\widetilde{p}^{\Delta t}_j(x,dy) = (\widetilde{X}^x_{j\Delta t})_* P$.

6.4.1. EM error bounds. We end this section by specializing the results to the EM method, $\widetilde{b}_{\Delta t}\equiv b$.

If we assume b is C^1 with bounded first derivative and Db is L-Lipschitz, then Taylor expanding b gives

$$\begin{aligned} \int_0^{\Delta t} E_P\left[\|\widetilde{b}_{\Delta t}(y)-b(\widetilde{X}^y_s)\|^2\right]ds \leq& \operatorname{tr}\left(Db(y)Db(y)^T\right)\frac{\Delta t^2}{2}\\ &+\|Db(y)b(y)\|^2\frac{\Delta t^3}{3}+\frac{16\sqrt{2}\Gamma((n+3)/2)}{5\Gamma(n/2)}L\|Db\|_\infty\Delta t^{5/2}\\ &+\frac{2n(n+2)L^2}{3}\Delta t^3+L\|Db\|_\infty\|b(y)\|^3\Delta t^4+\frac{2L^2}{5}\|b(y)\|^4\Delta t^5, \end{aligned} \tag{6.30}$$

and therefore

$$\begin{aligned} H_{N\Delta t}\leq& \frac{1}{N\Delta t}R(\widetilde{\mu}||\mu)+\frac{\Delta t}{4}\frac{1}{N}\sum_{j=1}^N\int E_P\left[\|Db(\widetilde{X}^x_{(j-1)\Delta t})\|_F^2\right]\widetilde{\mu}(dx)\\ &+\Delta t^{3/2}\Bigg(\frac{8\sqrt{2}\Gamma((n+3)/2)}{5\Gamma(n/2)}L\|Db\|_\infty+\frac{n(n+2)L^2}{3}\Delta t^{1/2}\\ &+\frac{1}{N}\sum_{j=1}^N\int E_P\Bigg[\frac{\Delta t^{1/2}}{6}\|Db(\widetilde{X}^x_{(j-1)\Delta t})b(\widetilde{X}^x_{(j-1)\Delta t})\|^2\\ &+\frac{L\|Db\|_\infty}{2}\|b(\widetilde{X}^x_{(j-1)\Delta t})\|^3\Delta t^{3/2}+\frac{L^2}{5}\|b(\widetilde{X}^x_{(j-1)\Delta t})\|^4\Delta t^{5/2}\Bigg]\widetilde{\mu}(dx)\Bigg), \end{aligned} \tag{6.31}$$

where $\|\cdot\|_F$ denotes the Frobenius matrix norm.

This is not the tightest possible bound, and alternatives can be obtained by Taylor expanding further, but it gives an idea of the type of result that can be obtained under various smoothness assumptions on b.

If the initial distributions have the form $d\widetilde{\mu} = e^{-\widetilde{\phi}}dx$ and $d\mu = e^{-\phi}dx$, where $\widetilde{\phi}$ and ϕ are known functions, then the relative entropy term takes the form

$$R(\widetilde{\mu}||\mu) = \int (\phi(x) - \widetilde{\phi}(x))e^{-\widetilde{\phi}(x)}dx. \tag{6.32}$$

If one can efficiently sample from $\widetilde{\mu}$, then (6.31) and (6.32) can be estimated via Monte Carlo methods, providing UQ bounds that involve a mixture of a priori and a posteriori data.

Appendix A. Proof of the perturbation bound.

Lemma A.1. *Let H be a Hilbert space, $A : D(A) \subset H \to H$ be a linear operator, and $B : H \to H$ be a bounded self-adjoint operator. Suppose there exist $D > 0$ and $x_0 \in H$ with $\|x_0\| = 1$ such that*

$$\langle Bx_0, x_0\rangle = 0 \quad \textit{and} \quad \mathrm{Re}(\langle Ax, x\rangle) \leq -D\|P^{\perp}x\|^2 \tag{A.1}$$

for all $x \in D(A)$, where $P^{\perp}$ is the orthogonal projector onto $x_0^{\perp}$.

Define

$$B^+ \equiv \max\left\{\sup_{\|y\|=1}\langle By, y\rangle, 0\right\}. \tag{A.2}$$

Then for any $0 \leq c < D/B^+$ we have

$$\sup_{x\in D(A),\|x\|=1} \mathrm{Re}(\langle (A + cB)x, x\rangle) \leq \frac{c^2\|Bx_0\|^2}{D - cB^+}. \tag{A.3}$$

Proof. Let $x \in D(A)$ with $\|x\| = 1$. Define $a = \langle x_0, x\rangle$. (Here we will use the convention of linearity in the second argument.) We have $\|P^{\perp}x\|^2 = 1 - |a|^2$, and so $|a| \leq 1$ with equality if and only if $P^{\perp}x = 0$.

We can decompose $x = ax_0 + \sqrt{1 - |a|^2}v$, where either $v = 0$ and $|a| = 1$ if $P^{\perp}x = 0$ or $v = P^{\perp}x/\sqrt{1 - |a|^2}$ and $\|v\| = 1$ if $P^{\perp}x \neq 0$. In either case, $v \perp x_0$.

With this, we have

$$\begin{aligned}
&\sup_{x\in D(A),\|x\|=1} \mathrm{Re}(\langle (A + cB)x, x\rangle) = \sup_{x\in D(A),\|x\|=1} \{\mathrm{Re}(\langle Ax, x\rangle) + c\,\mathrm{Re}(\langle Bx, x\rangle)\} \\
&\quad\leq \sup_{\beta\in[0,1]}\left\{-D(1 - \beta^2) + 2c\,\mathrm{Re}(\langle\sqrt{1 - \beta^2}v, aBx_0\rangle) + c(1 - \beta^2)\langle Bv, v\rangle\right\} \\
&\quad\leq \sup_{\beta\in[0,1]}\left\{2c\beta\sqrt{1 - \beta^2}\|Bx_0\| - \left(D - cB^+\right)(1 - \beta^2)\right\},
\end{aligned} \tag{A.4}$$

where B^+ is given by (A.2).

Restricting to $0 \leq c < D/B^+$, if $\|Bx_0\| = 0$, then the supremum is 0, and we have the result. Otherwise, the supremum is positive, and we can use $\beta \leq 1/\beta$ and then change variables in the supremum to $r=\sqrt{1-\beta^2}/\beta$, thereby obtaining

$$\begin{aligned}\text{(A.5)}\qquad &\sup_{x\in D(A),\|x\|=1} \operatorname{Re}(\langle (A+cB)x,x\rangle)\\ &\leq \sup_{\beta\in(0,1]}\left\{\beta\sqrt{1-\beta^2}\left(2c\|Bx_0\|-\left(D-cB^+\right)\sqrt{1-\beta^2}/\beta\right)\right\}\\ &\leq \sup_{r\geq 0}\left\{2c\|Bx_0\|r-\left(D-cB^+\right)r^2\right\}=\frac{\|Bx_0\|^2c^2}{D-cB^+}.\end{aligned}$$

■

The previous lemma is closest in spirit to the probabilistic application, as $\|Bx_0\|^2$ plays the role of the variance. However, one can work with non-self-adjoint perturbations if one instead uses the definition

$$B^+\equiv\max\left\{\sup_{\|y\|=1}\operatorname{Re}(\langle By,y\rangle),0\right\} \tag{A.6}$$

and makes the replacement $\|Bx_0\|\to\|(B+B^*)x_0/2\|$ in (A.3). The proof is similar.

Appendix B. F-Sobolev inequalities.

Proposition 4.12 can be generalized to the F-Sobolev case; see the proof of Theorem 2.3 in [12].

Proposition B.1. *Let A be the generator of $\mathcal{P}_t$ and μ^* be an invariant measure. Suppose we have a function $F:(0,\infty)\to\mathbb{R}$ satisfying the following:*

1. *F is strictly increasing,*
2. *F is concave (hence continuous),*
3. *$F(1)=0$,*
4. *$F(x)\to\infty$ as $x\to\infty$, and*
5. *$F(xy)\leq F(x)+F(y)$ for all $x,y\geq 0$.*

(Note that this implies $F^{-1}:(F(0^+),\infty)\to(0,\infty)$ exists and is increasing, convex, and continuous.)

Assume the F-Sobolev inequality holds for μ^*:

$$\int g^2F(g^2)d\mu^*\leq-\int A[g]gd\mu^*\quad \text{for all } g\in D(A,\mathbb{R}) \text{ with } \|g\|_{L^2(\mu^*)}=1. \tag{B.1}$$

Finally, suppose that $V\in L^1(\mu^*)$ with $V>F(0^+)$ and $\int F^{-1}(V)d\mu^*<\infty$. Then $\mathcal{P}_t^V: L^2(\mu^*)\to L^2(\mu^*)$, defined by

$$\mathcal{P}_t^V[g](x)=E^x\left[g(X_t)\exp\left(\int_0^t V(X_s)ds\right)\right], \tag{B.2}$$

are well-defined linear operators, and the operator norm satisfies the bound

$$\|\mathcal{P}_t^V\|\leq\exp\left[tF\left(\int F^{-1}(V)d\mu^*\right)\right]. \tag{B.3}$$

Note that if $F(0^+)=-\infty$, then certain unbounded observables are allowed, namely those that satisfy the integrability condition (4.38).

This proposition leads to a UQ bound of the form (2.23). The proof is analogous to the log-Sobolev case from section 4.4.

Theorem B.2. *In addition to Assumption 2.8, assume the F-Sobolev inequality, (B.1), holds for some function, F, having the properties listed in Proposition B.1, $f \in L^1(\mu^*, \mathbb{R})$, and there exists $c_- < 0 < c_+$ such that, for all $c \in (c_-, c_+)$,*

$$F(0^+) < \pm c(f - \mu^*[f]), \quad \int F^{-1}\left(\pm c\left(f - \mu^*[f]\right)\right) d\mu^* < \infty. \tag{B.4}$$

Then a UQ bound of the form (2.23) holds with

$$\Lambda(c) = \begin{cases} F\left(\int F^{-1}(V_c) d\mu^*\right) & \text{if } c \in (c_-, c_+), \\ +\infty & \text{otherwise.} \end{cases} \tag{B.5}$$

In addition, if $\mathrm{Var}_{\mu^*}[f] > 0$, F and F^{-1} are smooth, $F'(1) > 0$, $(F^{-1})''(0) > 0$, and $c \to \mu^*[F^{-1}(V_c)]$ is smooth on a neighborhood of 0 and can be differentiated under the integral, then (2.16) holds with

$$\Lambda''(0) = F'(1)\,(F^{-1})''(0)\,\mathrm{Var}_{\mu^*}[f], \quad \eta = \frac{1}{T} R(\widetilde{P}_T^{\widetilde{\mu}} || P_T^{\mu^*}). \tag{B.6}$$

Appendix C. Continuous-time jump processes on general state spaces.

As discussed in section 5, to apply our UQ bounds to the invariant measure of discrete-time Markov processes, $\mathcal{P}$ and $\widetilde{P}$, one needs to construct an ancillary continuous-time Markov process with generators $\mathcal{P} - I$ and $\widetilde{\mathcal{P}} - I$ and also compute the associated relative entropy. While the construction of continuous-time Markov processes from their generators is well known (see, for example, Chapter 4.2 in [26] or Chapter 3.3 in [47]), and the relative entropy computation is known in the countable state space case (see section 6.1), we require a formula for the relative entropy in the general case of a Polish state space. To the best of our knowledge, this computation is new, though it closely mirrors the established results; hence we present only a short outline.

In order to obtain an explicit formula for the Radon–Nikodym derivative, and thereby compute the relative entropy, it is useful to utilize an explicit construction, as in the countable state space case (see, for example, Appendix 1 of [42]), rather than invoking more general existence theorems.

Let $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ be a Polish space and $p(x, dy)$ be a probability kernel on $\mathcal{X}$. Given $\lambda > 0$, define the probability kernel, p^J, on the Polish space $(\mathcal{X} \times (0, \infty), \mathcal{B}_{\mathcal{X}} \bigotimes \mathcal{B}_{(0,\infty)})$:

$$p^J((x, s), \cdot) = p(x, dy) \times \lambda e^{-\lambda t} dt. \tag{C.1}$$

For any probability measure π on $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$, let P^π (for $\pi = \delta_x$ we simply write P^x) be the unique probability measure on $(\Omega, \mathcal{F}) \equiv (\prod_{n=0}^\infty (\mathcal{X} \times (0, \infty)), \bigotimes_{n=0}^\infty (\mathcal{B}_{\mathcal{X}} \bigotimes \mathcal{B}_{(0,\infty)}))$ generated by the transition probabilities p^J and initial distribution $\pi \times (\lambda e^{-\lambda t} dt)$. Also, define the jump process, jump intervals, and jump times:

$$X_n^J \equiv \pi_1 \circ \pi_n, \;\; \Delta_n \equiv \mu \circ \pi_n \text{ for } n \in \mathbb{Z}_0, \qquad J_0 \equiv 0, \;\; J_n \equiv \sum_{k=0}^{n-1} \Delta_k \text{ for } n \in \mathbb{Z}^+, \tag{C.2}$$

where π_i denote projections onto components. The jump rates are positive constants, so one obtains $J_n(\omega) \to \infty$ a.s. as $n \to \infty$.

(X_n^J, Δ_n) is a Markov process under P^π with transition probabilities p^J and initial distribution $\pi \times (\lambda e^{\lambda t} dt)$. Use this to define the associated càdlàg process

$$X_t(\omega) = X_n^J, \quad \text{where } t \in [J_n(\omega), J_{n+1}(\omega)), \tag{C.3}$$

and the probability kernels on $\mathcal{X}$,

$$p_t(x, A) \equiv P^x(X_t \in A), \quad t \geq 0, \ x \in \mathcal{X}. \tag{C.4}$$

Finally, let $\mathcal{F}_t$ be the natural filtration for X_t.

With this setup, we have the following theorem.

Theorem C.1. *$(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, P^x)$, $x \in \mathcal{X}$, is a càdlàg Markov family with transition probabilities p_t. More specifically, the following hold:*

1. *$(\Omega, \mathcal{F}, \mathcal{F}_t)$, $t \geq 0$, is a filtered probability space and X_t is an $\mathcal{X}$-valued, $\mathcal{F}_t$-adapted, càdlàg process.*
2. *$p_t(x, dy)$, $t \geq 0$, are time homogeneous transition probabilities on $\mathcal{X}$.*
3. *P^x, $x \in \mathcal{X}$, are probability measures with $(X_0)_* P^x = \delta_x$ for each $x \in \mathcal{X}$.*
4. *For every measurable set F, $x \to P^x(F)$ is universally measurable.*
5. *For each $x \in \mathcal{X}$, $P^x(X_{t+s} \in B | \mathcal{F}_s) = p_t(X_s, B)$ P^x-a.s. In particular, $p_t(x, B) = P^x(X_t \in B)$.*

One also obtains realizability of the semigroup $\exp(t\lambda(\mathcal{P} - I))$ by a probability kernel.

Theorem C.2. *If μ^* is an invariant measure for p, then μ^* is invariant for p_t for all $t \geq 0$ and the bounded linear operators on $L^2(\mu^*)$,*

$$\mathcal{P}[f](x) \equiv \int f(y) p(x, dy), \quad \mathcal{P}_t[f](x) \equiv \int f(y) p_t(x, dy), \tag{C.5}$$

satisfy

$$\mathcal{P}_t = \exp(t\lambda(\mathcal{P} - I)) \tag{C.6}$$

for all $t \geq 0$, where the right-hand side is the operator exponential for bounded operators on $L^2(\mu^*)$.

These results are all straightforward to prove by using the same strategy as the discrete state space case.

The formula for the Radon–Nikodym derivative for two measures constructed as above is also straightforward; the only complication is that here, the jump chain (X_n^J, Δ_n) is generally not recoverable from X_t; specifically, the J_n are not $\mathcal{F}_t$-stopping times (this is because "jumps" do not necessarily change the state, unlike the construction commonly used when the state space is discrete). Hence, we must derive a formula for the Radon–Nikodym derivative on the enlarged filtration

$$\mathcal{G}_t \equiv \sigma(1_{J_n \leq s}, X_{J_n \wedge s} : s \leq t, n \geq 0). \tag{C.7}$$

Otherwise, the computation closely mirrors the discrete case (again, see [42]), and one arrives at the following theorem.

Theorem C.3. *Suppose we have probability measures $\widetilde{\mu}$, μ and probability kernels $\widetilde{p}(x,dy)$, $p(x,dy)$ on $\mathcal{X}$. Assume that $\widetilde{\mu} \ll \mu$ and $\widetilde{p}(x,\cdot) \ll p(x,\cdot)$ for $\widetilde{\mu}$ a.e. x. In particular, we have $h \in L^+(\mathcal{X}\times\mathcal{X})$ such that*

$$\widetilde{p}(x,dy) = h(x,y)p(x,dy) \text{ for } \widetilde{\mu} \text{ a.e. } x. \tag{C.8}$$

Given $\lambda > 0$, construct the probability measures $\widetilde{P}^{\widetilde{\mu}}$ and P^μ on Ω from $\widetilde{p}$ and p, respectively, and define the process X_t as in (C.3).

Suppose $(\widetilde{\mathcal{P}}^\dagger)^n[\widetilde{\mu}] \ll \widetilde{\mu}$ for all n (in particular, if $\widetilde{\mu}$ is invariant for $\widetilde{p}$). Then for any $t \geq 0$ we have $\widetilde{P}^{\widetilde{\mu}}|_{\mathcal{G}_t} \ll P^\mu|_{\mathcal{G}_t}$ and

$$\frac{d\widetilde{P}^{\widetilde{\mu}}|_{\mathcal{G}_t}}{dP^\mu|_{\mathcal{G}_t}} = \frac{d\widetilde{\mu}}{d\mu}(X_0) \prod_{n\geq 1: J_n \leq t} h(X_{J_{n-1}\wedge t}, X_{J_n \wedge t}). \tag{C.9}$$

By an analogous computation to the continuous-time Markov chain case, (6.2), the formula for the Radon–Nikodym derivative (C.9) leads to the following formula for the relative entropy.

Theorem C.4. *Suppose we have probability measures $\widetilde{\mu}$, μ and probability kernels $\widetilde{p}(x,dy)$, $p(x,dy)$ on $\mathcal{X}$. Assume that $\widetilde{\mu} \ll \mu$ and $\widetilde{p}(x,\cdot) \ll p(x,\cdot)$ for $\widetilde{\mu}$ a.e. x.*

Suppose $(\widetilde{\mathcal{P}}^\dagger)^n[\widetilde{\mu}] \ll \widetilde{\mu}$ for all n (in particular, if $\widetilde{\mu}$ is invariant for $\widetilde{p}$). Then for any $t \geq 0$ we have

$$R(\widetilde{P}^{\widetilde{\mu}}|_{\mathcal{G}_t}||P^\mu|_{\mathcal{G}_t}) = R(\widetilde{\mu}||\mu) + \lambda \int_0^t \widetilde{E}^{\widetilde{\mu}}\left[\int \log(h(X_s,z))h(X_s,z)p(X_s,dz)\right] ds, \tag{C.10}$$

where h is as defined in (C.8).

It is also useful to note that, by the data processing inequality (see Theorem 14 in [46]), $\mathcal{F}_t \subset \mathcal{G}_t$ implies

$$R(\widetilde{P}^{\widetilde{\mu}}|_{\mathcal{F}_t}||P^\mu|_{\mathcal{F}_t}) \leq R(\widetilde{P}^{\widetilde{\mu}}|_{\mathcal{G}_t}||P^\mu|_{\mathcal{G}_t}). \tag{C.11}$$

When $\widetilde{\mu}$ is an invariant measure we obtain the following simpler formula.

Corollary C.5. *Suppose we have probability measures $\widetilde{\mu}^*$, μ and probability kernels $\widetilde{p}(x,dy)$, $p(x,dy)$ on $\mathcal{X}$. If $\widetilde{\mu}^*$ is invariant for $\widetilde{p}$, then for all $t > 0$*

$$R(\widetilde{P}^{\widetilde{\mu}^*}|_{\mathcal{F}_t}||P^\mu|_{\mathcal{F}_t}) \leq R(\widetilde{\mu}^*||\mu) + \lambda t \int R(\widetilde{p}(x,\cdot)||p(x,\cdot))d\widetilde{\mu}^*. \tag{C.12}$$

This is the relative entropy bound that was used in section 5 when applying our UQ results to invariant measures of discrete-time Markov processes.

REFERENCES

[1] P. Alquier, R. Friel, R. Everitt, and A. Boland, *Noisy Monte Carlo: Convergence of Markov chains with approximate transition kernels*, Stat. Comput., 26 (2016), pp. 29–47, https://doi.org/10.1007/s11222-014-9521-x.

[2] D. Bakry, F. Barthe, P. Cattiaux, and A. Guillin, *A simple proof of the Poincaré inequality for a large class of probability measures*, Electron. Commun. Probab., 13 (2008), pp. 60–66, https://doi.org/10.1214/ECP.v13-1352.
[3] D. Bakry, P. Cattiaux, and A. Guillin, *Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré*, J. Funct. Anal., 254 (2008), pp. 727–759, https://doi.org/10.1016/j.jfa.2007.11.002.
[4] D. Bakry and M. Emery, *Hypercontractivité de semi-groupes de diffusion*, C. R. Acad. Sci. Paris Sér. I Math., 299 (1984), pp. 775–778.
[5] R. Bardenet, A. Doucet, and C. Holmes, *Towards scaling up Markov chain Monte Carlo: An adaptive subsampling approach*, in Proceedings of the 31st International Conference on Machine Learning (Bejing, China, 2014), E. P. Xing and T. Jebara, eds., Proc. Machine Learn. Res. 32, pp. 405–413, http://proceedings.mlr.press/v32/bardenet14.html.
[6] J. Birrell and L. Rey-Bellet, *Concentration Inequalities and Performance Guarantees for Hypocoercive MCMC Samplers*, preprint, https://arxiv.org/abs/1907.11973, 2019.
[7] M. Bladt and B. Nielsen, *Matrix-Exponential Distributions in Applied Probability*, Probab. Theory Stochastic Model. 81, Springer, New York, 2017, https://doi.org/10.1007/978-1-4939-7049-0.
[8] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press, Oxford, UK, 2013, https://doi.org/10.1093/acprof:oso/9780199535255.001.0001.
[9] T. Breuer and I. Csiszár, *Measuring distribution model risk*, Math. Finance, 26 (2016), pp. 395–411, https://doi.org/10.1111/mafi.12050.
[10] T. Breuer and I. Csiszár, *Systematic stress tests with entropic plausibility constraints*, J. Banking & Finance, 37 (2013), pp. 1552–1559, https://doi.org/10.1016/j.jbankfin.2012.04.013.
[11] E. Carlen and M. Loss, *Logarithmic Sobolev inequalities and spectral gaps*, in Recent Advances in the Theory and Applications of Mass Transport, Contemp. Math. 353, AMS, Providence, RI, 2004, pp. 53–60.
[12] P. Cattiaux and A. Guillin, *Deviation bounds for additive functionals of Markov processes*, ESAIM Probab. Stat., 12 (2008), pp. 12–29, https://doi.org/10.1051/ps:2007032.
[13] P. Cattiaux, A. Guillin, and C. Roberto, *Poincaré inequality and the L^p convergence of semi-groups*, Electron. Commun. Probab., 15 (2010), pp. 270–280, https://doi.org/10.1214/ECP.v15-1559.
[14] D. Chafaï, *Binomial-Poisson entropic inequalities and the $M/M/\infty$ queue*, ESAIM Probab. Stat., 10 (2006), pp. 317–339, https://doi.org/10.1051/ps:2006013.
[15] K. Chowdhary and P. Dupuis, *Distinguishing and integrating aleatoric and epistemic variation in uncertainty quantification*, ESAIM Math. Model. Numer. Anal., 47 (2013), pp. 635–662, https://doi.org/10.1051/m2an/2012038.
[16] G. Deligiannidis, D. Paulin, A. Bouchard-Côté, and A. Doucet, *Randomized Hamiltonian Monte Carlo as Scaling Limit of the Bouncy Particle Sampler and Dimension-Free Convergence Rates*, preprint, https://arxiv.org/abs/1808.04299, 2018.
[17] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Stochastic Model. Appl. Probab. 38, Springer, Berlin, Heidelberg, 2010, https://doi.org/10.1007/978-3-642-03311-7.
[18] P. Diaconis and L. Saloff-Coste, *Comparison theorems for reversible Markov chains*, Ann. Appl. Probab., 3 (1993), pp. 696–730, https://doi.org/10.1214/aoap/1177005359.
[19] P. Diaconis and L. Saloff-Coste, *Logarithmic Sobolev inequalities for finite Markov chains*, Ann. Appl. Probab., 6 (1996), pp. 695–750, https://doi.org/10.1214/aoap/1034968224.
[20] J. Dolbeault, C. Mouhot, and C. Schmeiser, *Hypocoercivity for kinetic equations with linear relaxation terms*, C. R. Math. Acad. Sci. Paris, 347 (2009), pp. 511–516, https://doi.org/10.1016/j.crma.2009.02.025.
[21] J. Dolbeault, C. Mouhot, and C. Schmeiser, *Hypocoercivity for linear kinetic equations conserving mass*, Trans. Amer. Math. Soc., 367 (2015), pp. 3807–3828, https://doi.org/10.1090/S0002-9947-2015-06012-7.
[22] P. Dupuis and R. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*, Wiley Ser. Probab. Stat., Wiley, New York, 2011, https://doi.org/10.1002/9781118165904.
[23] P. Dupuis, M. A. Katsoulakis, Y. Pantazis, and P. Plecháč, *Path-space information bounds for uncertainty quantification and sensitivity analysis of stochastic dynamics*, SIAM/ASA J. Uncertain.

Quantif., 4 (2016), pp. 80–111, https://doi.org/10.1137/15M1025645.
[24] P. Dupuis, M. A. Katsoulakis, Y. Pantazis, and L. Rey-Bellet, *Sensitivity Analysis for Rare Events Based on Rényi Divergence*, preprint, https://arxiv.org/abs/1805.06917, 2018.
[25] K.-J. Engel and R. Nagel, *A Short Course on Operator Semigroups*, Universitext, Springer, New York, 2006, https://doi.org/10.1007/0-387-36619-9.
[26] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence*, Wiley Ser. Probab. Stat., Wiley, New York, 2005, https://doi.org/10.1002/9780470316658.
[27] M. J. Faddy, *Examples of fitting structured phase-type distributions*, Appl. Stochastic Models Data Anal., 10 (1994), pp. 247–255, https://doi.org/10.1002/asm.3150100403.
[28] D. Ferré, L. Hervé, and J. Ledoux, *Regular perturbation of* V*-geometrically ergodic Markov chains*, J. Appl. Probab., 50 (2013), pp. 184–194, https://doi.org/10.1239/jap/1363784432.
[29] F. Gao, A. Guillin, and L. Wu, *Bernstein-type concentration inequalities for symmetric Markov processes*, Theory Probab. Appl., 58 (2014), pp. 358–382, https://doi.org/10.1137/S0040585X97986667.
[30] V. Girardin and N. Limnios, *On the entropy for semi-Markov processes*, J. Appl. Probab., 40 (2003), pp. 1060–1068, https://doi.org/10.1239/jap/1067436100.
[31] P. Glasserman and X. Xu, *Robust risk measurement and model risk*, Quant. Finance, 14 (2014), pp. 29–58, https://doi.org/10.1080/14697688.2013.822989.
[32] K. Gourgoulias, M. A. Katsoulakis, and L. Rey-Bellet, *Information metrics for long-time errors in splitting schemes for stochastic dynamics and parallel kinetic Monte Carlo*, SIAM J. Sci. Comput., 38 (2016), pp. A3808–A3832, https://doi.org/10.1137/15M1047271.
[33] K. Gourgoulias, M. A. Katsoulakis, L. Rey-Bellet, and J. Wang, *How Biased Is Your Model? Concentration Inequalities, Information and Model Bias*, preprint, https://arxiv.org/abs/1706.10260, 2017.
[34] A. Guillin, C. Léonard, L. Wu, and N. Yao, *Transportation-information inequalities for Markov processes*, Probab. Theory Related Fields, 144 (2009), pp. 669–695, https://doi.org/10.1007/s00440-008-0159-5.
[35] M. Hairer, *Convergence of Markov Processes*, Lecture notes, 2010.
[36] M. Hairer and J. Mattingly, *Yet another look at Harris' ergodic theorem for Markov chains*, in Seminar on Stochastic Analysis, Random Fields and Applications VI, Progr. Probab. 63, Birkhäuser, Basel, 2011, pp. 109–117.
[37] J. H. Huggins and J. Zou, *Quantifying the Accuracy of Approximate Diffusions and Markov Chains*, preprint, https://arxiv.org/abs/1605.06420, 2016.
[38] J. Janssen and R. Manca, *Applied Semi-Markov Processes*, Springer, New York, 2006, https://doi.org/10.1007/0-387-29548-8.
[39] J. E. Johndrow, J. C. Mattingly, S. Mukherjee, and D. Dunson, *Optimal Approximating Markov Chains for Bayesian Inference*, preprint, https://arxiv.org/abs/1508.03387, 2015.
[40] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed., Grad. Texts in Math. 113, Springer, New York, 1998, https:doi.org/10.1007/978-1-4612-0949-2.
[41] M. A. Katsoulakis, L. Rey-Bellet, and J. Wang, *Scalable information inequalities for uncertainty quantification*, J. Comput. Phys., 336 (2017), pp. 513–545, https://doi.org/10.1016/j.jcp.2017.02.020.
[42] C. Kipnis and C. Landim, *Scaling Limits of Interacting Particle Systems*, Grundlehren Math. Wiss. 320, Springer, Berlin, Heidelberg, 1999, https://doi.org/10.1007/978-3-662-03752-2.
[43] H. Lam, *Robust sensitivity analysis for stochastic systems*, Math. Oper. Res., 41 (2016), pp. 1248–1275, https://doi.org/10.1287/moor.2015.0776.
[44] D. A. Levin and Y. Peres, *Markov Chains and Mixing Times*, 2nd ed., AMS, Providence, RI, 2017, https://doi.org/10.1090/mbk/107.
[45] P. Lezaud, *Chernoff and Berry-Esséen inequalities for Markov processes*, ESAIM Probab. Stat., 5 (2001), pp. 183–201, https://doi.org/10.1051/ps:2001108.
[46] F. Liese and I. Vajda, *On divergences and informations in statistics and information theory*, IEEE Trans. Inform. Theory, 52 (2006), pp. 4394–4412, https://doi.org/10.1109/TIT.2006.881731.
[47] T. Liggett, *Continuous Time Markov Processes: An Introduction*, Grad. Stud. Math. 113, AMS, Providence, RI, 2010, https://doi.org/10.1090/gsm/113.
[48] N. Limnios and G. Oprişan, *Semi-Markov Processes and Reliability*, Stat. Ind. Technol., Birkhäuser Boston, Boston, 2001, https://doi.org/10.1007/978-1-4612-0161-8.

[49] F. J. Medina-Aguayo, A. Lee, and G. O. Roberts, *Stability of noisy Metropolis–Hastings*, Stat. Comput., 26 (2016), pp. 1187–1211, https://doi.org/10.1007/s11222-015-9604-3.
[50] A. Mitrophanov, *Sensitivity and convergence of uniformly ergodic Markov chains*, J. Appl. Probab., 42 (2005), pp. 1003–1014, https://doi.org/10.1239/jap/1134587812.
[51] Y. Pantazis and M. Katsoulakis, *A relative entropy rate method for path space sensitivity analysis of stationary complex stochastic dynamics*, J. Chem. Phys., 138 (2013), 054115, https://doi.org/10.1063/1.4789612.
[52] N. S. Pillai and A. Smith, *Ergodicity of Approximate MCMC Chains with Applications to Large Data Sets*, preprint, https://arxiv.org/abs/1405.0182, 2014.
[53] O. S. Rothaus, *Lower bounds for eigenvalues of regular Sturm-Liouville operators and the logarithmic Sobolev inequality*, Duke Math. J., 45 (1978), pp. 351–362, https://doi.org/10.1215/S0012-7094-78-04518-0.
[54] D. Rudolf and N. Schweizer, *Perturbation theory for Markov chains via Wasserstein distance*, Bernoulli, 24 (2018), pp. 2610–2639, https://doi.org/10.3150/17-BEJ938.
[55] F. Wang, *Functional Inequalities Markov Semigroups and Spectral Theory*, The Science Series of the Contemporary Elite Youth, Elsevier Science, Amsterdam, 2006.
[56] L. Wu, *A deviation inequality for non-reversible Markov processes*, Ann. Inst. H. Poincaré Probab. Statist., 36 (2000), pp. 435–445, https://doi.org/10.1016/S0246-0203(00)00135-7.
[57] K. Yosida, *Functional Analysis*, 3rd ed., Springer, New York, 1971.