QUARTERLY OF APPLIED MATHEMATICS
VOLUME LXXXIV, NUMBER 4
DECEMBER 2026, PAGES 565–583
https://doi.org/10.1090/qam/1732
Article electronically published on November 14, 2025

CONCENTRATION INEQUALITIES AND UQ BOUNDS FOR HYPOCOERCIVE MCMC SAMPLERS

By

JEREMIAH BIRRELL (*Department of Mathematics, Texas State University, 601 University Drive, San Marcos, TX 78666*)

AND

LUC REY-BELLET (*Department of Mathematics and Statistics, University of Massachusetts Amherst, 710 N Pleasant Street, Amherst, MA 01002*)

Dedicated to Professor Paul Dupuis on the occasion of his 65th birthday

Abstract. In this work we provide performance guarantees for hypocoercive nonreversible Monte-Carlo Markov chain (MCMC) samplers X_t with invariant measure μ_*; our results apply in particular to the Langevin equation, Hamiltonian Monte-Carlo, and the bouncy particle and zig-zag samplers. Specifically, we establish a concentration inequality of Bernstein type for ergodic averages $\frac{1}{T}\int_0^T f(X_t)\,dt$. As a consequence we provide two types of performance guarantees: (a) explicit nonasymptotic confidence intervals for $\int f d\mu_*$ when using a finite time ergodic average with given initial condition μ and (b) uncertainty quantification (UQ) bounds, expressed in terms of relative entropy rate, on the bias of $\int f d\mu_*$ when using an alternative or approximate processes $\widetilde{X}_t$. (Results in (b) generalize results of the authors [*Uncertainty quantification for Markov processes via variational principles and functional inequalities*, SIAM/ASA J. Uncertain. Quantif. 8 (2020), no. 2, 539–572] for coercive dynamics.) The concentration inequality is proved by combining the approach via Feynman-Kac semigroups first noted by Wu [*A deviation inequality for non-reversible Markov processes*, Ann. Inst. H. Poincaré Probab. Statist. 36 (2000), no. 4, 435–445] with the hypocoercive estimates of Dolbeault, Mouhot, and Schmeiser [*Hypocoercivity for kinetic equations with linear relaxation terms*, C. R. Math. Acad. Sci. Paris 347 (2009), no. 9–10, 511–516] and [*Hypocoercivity for linear kinetic equations conserving mass*, Trans. Amer. Math. Soc. 367 (2015), no. 6, 3807–3828] developed for the Langevin equation and generalized to partially deterministic Markov

Received June 23, 2025, and, in revised form, October 9, 2025.
2020 *Mathematics Subject Classification.* Primary 60J25, 47D07, 39B72.
Research supported in part by the National Science Foundation (DMS-1515712, DMS-2008970) and the Air Force Office of Scientific Research (AFOSR) (FA-9550-18-1-0214).
Email address: jbirrell@txstate.edu
Email address: luc@math.umass.edu

©2025 Brown University

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

processes by Andrieu, Durmus, Nüsken, and Roussel [*Hypocoercivity of piecewise deterministic Markov process-Monte Carlo*, Ann. Appl. Probab. 31 (2021), no. 5, 2478–2517].

1. Introduction and statement of the results. Consider the problem of computing the expected value

$$\nu_*[f] \equiv \int f(q) d\nu_*(q) \quad \text{with } d\nu_* = Z^{-1} e^{-\beta V(q)} dq \tag{1.1}$$

for some given function $f : \mathcal{Q} \to \mathbb{R}$ with $\mathcal{Q} \subset \mathbb{R}^d$. If the normalization constant Z is unknown or prohibitive to compute, it can be advantageous to construct an ergodic stochastic process Q_t with stationary distribution ν_* (in this paper, only continuous-time processes are considered) and use the fact that, by the strong law of large numbers and for suitable f, one has

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(Q_t)\, dt = \nu_*[f]. \tag{1.2}$$

Such a process Q_t is usually called a Monte-Carlo Markov chain (MCMC) and we can then use the finite time average $\frac{1}{T}\int_0^T f(Q_t)dt$ as an estimator for $\nu_*[f]$. There are of course multiple choices of stochastic processes with invariant measure ν_* and in order to decide which process to use we need to evaluate its performance.

While traditional Monte-Carlo algorithms are often built to be reversible, in recent years nonreversible algorithms have attracted a lot of attention because of their potential to sample the space in a more efficient manner, in particular in the context of Bayesian statistics and molecular dynamics (see for example [3, 14, 18, 33, 43] and many more references therein). In this paper we consider a variety of nonreversible MCMC samplers such as the Langevin equation, and various modifications thereof, as well as partially deterministic Markov processes such as the zig-zag sampler [4], the bouncy particle sampler [40], and the hybrid Hamiltonian Monte-Carlo [17]. Each of these samplers are constructed by extending the phase space from $\mathcal{Q}$ to $\mathcal{X} = \mathcal{Q} \times \mathcal{P}$ and then constructing a nonreversible MCMC in the extended phase space with an invariant measure $\mu_* = \nu_* \times \rho_*$. The extra dimension $\mathcal{P}$ can be thought of as momentum space and the dynamics considered here combine a conservative Hamiltonian-type dynamics with a dissipative sampling mechanism for the measure ρ_* in the momentum variable $p \in \mathcal{P}$. For example, ρ_* may be a Gaussian distribution, although other choices are possible; we do assume ρ_* has mean zero and nonzero covariance.

All of the algorithms we consider here have been proved to be hypocoercive. The concept of hypocoercivity was formalized by Villani to describe dynamics which do not satisfy a Poincaré inequality (otherwise they would be called coercive) but yet converge exponentially fast to equilibrium in $L^2(\mu_*)$. In a series of work [13, 24, 31, 32, 47] it was proved that the Langevin equation is hypocoercive (see also [23, 25, 37, 44] for some earlier and related convergence results). More recently [15, 16] found a new, short, and very elegant proof of hypocoercivity, and their techniques have been used for various modifications of the Langevin equation [34, 45, 46] (some of them without hypoellipticity)

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

and for a class of partially deterministic Markov processes [1], among them:

(1) The bouncy particle sampler, which was introduced in [40] and whose ergodic properties were studied in [9] and [48].
(2) The zig-zag sampler, introduced in [4] and further studied in [5], which generalizes to higher dimension the so-called telegraph process studied earlier in [26], [27], and [38].
(3) The hybrid Hamiltonian Monte Carlo introduced by [17] and whose ergodic properties are studied in [8]; see also [22] and [39].

To explain this result, decompose the generator A of the dynamics on the Hilbert space $L^2(\mu_*)$ into symmetric and antisymmetric parts, $A = S + T$ with $S^* = S$, $T^* = -T$, denote by Π the projection of $L^2(\mu_*)$ onto $L^2(\nu_*)$ given $\Pi(f)(q) = \int f(q,p)d\rho_*(p)$, and consider the operator

$$B = (I + (T\Pi)^*(T\Pi))^{-1}(-T\Pi)^* \tag{1.3}$$

and the family of modified scalar products $\langle f, g\rangle_\epsilon = \langle f, (I + \epsilon G)g\rangle$, where $G = B + B^*$; here, and in the following, $\langle\cdot,\cdot\rangle$ will denote the $L^2(\mu_*)$-scalar product. Under suitable conditions (more details are in Section 2) the norms induced by these modified scalar products are equivalent to the $L^2(\mu_*)$-norm and it is shown in [1, 15, 16] that, for sufficiently small values of $\epsilon > 0$, the dynamics satisfy a Poincaré inequality for the modified scalar product

$$\langle -Af, f\rangle_\epsilon \geq \Lambda(\epsilon)\ \mathrm{Var}_{\mu_*}(f), \tag{1.4}$$

where $\Lambda(\epsilon) > 0$ can be explicitly bounded in terms of the Poincaré constant of the measure ν_*, the spectral gap of the sampling dynamics for ρ_*, and properties of the potential V (see Section 2).

In this paper we leverage this approach to hypocoercivity to prove concentration inequalities (of Bernstein type) for finite time ergodic averages of a function (observable) $f : \mathcal{X} \to \mathbb{R}$:

$$F_T = \frac{1}{T}\int_0^T f(X_t)dt. \tag{1.5}$$

For reversible processes, or more generally for processes whose reversible part satisfies a Poincaré inequality, that is, for *coercive* processes, concentration inequalities were obtained first in [36] and then both simplified and greatly generalized in [11, 28, 30, 49]; our approach relies heavily on the ideas developed in those works.

Concentration inequalities are very useful for providing performance guarantees (e.g., confidence intervals) valid for all time T (and which do not rely on the central limit theorem). In practice, algorithm performance is typically evaluated in terms of the asymptotic variance

$$\sigma^2(f) \equiv \lim_{T\to\infty} T\,\mathrm{Var}\left(\frac{1}{T}\int_0^T f(X_t)\,dt\right); \tag{1.6}$$

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

we take here the alternative point of view of using concentration inequalities, thereby obtaining *nonasymptotic* performance guarantees. In a related manner, it has been advocated in [21] that, to evaluate the performance of an MCMC algorithm, one consider the large deviation rate for the empirical measure itself, an approach used in [42] to analyze nonreversible perturbations of the overdamped Langevin equation.

We prove the following nonasymptotic performance guarantee in Section 2 (see Corollary 3.3). Here, and in the following, (X_t, P^μ) will denote an $\mathcal{X}$-valued Markov process with initial distribution $X_0 \sim \mu$ (i.e., $\mu = (X_0)_* P^\mu$ is the pushforward of P^μ by X_0), E^μ will be the expectation with respect to P^μ, and $\|\cdot\|$ will denote the $L^2(\mu_*)$-norm.

THEOREM 1.1 (Nonasymptotic confidence intervals). Suppose that the Markov process (X_t, P^μ) satisfies the hypocoercive estimate (1.4). Then for any bounded observable f, any time $T > 0$, and tolerance level $0 < 1 - \delta < 1$ we have

$$P^\mu\left(\left|\frac{1}{T}\int_0^T f(X_t)dt - \mu_*[f]\right| \le r\right) \ge 1 - \delta, \tag{1.7}$$

where

$$r = \sqrt{2v\frac{1}{T}\log\left(\frac{2N}{\delta}\right)} + b\frac{1}{T}\log\left(\frac{2N}{\delta}\right), \tag{1.8}$$

with

$$v = \frac{(1+\epsilon)(1-\frac{\epsilon^2}{4})}{1-\epsilon}\frac{2\,\mathrm{Var}_{\mu_*}[f]}{\Lambda(\epsilon)}, \quad b = \frac{(1+\epsilon)^2}{1-\epsilon}\frac{\|\widehat{f}\|_\infty}{\Lambda(\epsilon)}, \quad N = \frac{\left\|\frac{d\mu}{d\mu_*}\right\|}{\sqrt{1-\epsilon}}. \tag{1.9}$$

$\Lambda(\epsilon)$ is defined in equation (2.22) and $\epsilon \in (0,1)$ must satisfy equation (2.23).

We emphasize that the novelty of Theorem 1.1 lies in our technique for obtaining concentration inequalities from a Poincaré inequality in a modified inner product, thereby extending established nonasymptotic results for coercive systems [11, 28, 30, 36, 49] to the hypocoercive setting for the first time. As our derivation fundamentally relies on the convergence rate obtained via the hypocoercivity method of [15, 16], we note that Theorem 1.1 does not provide substantial new insight regarding performance comparisons between coercive and hypocoercive samplers, beyond what is already apparent from the (modified) Poincaré inequalities. For the derivation of explicit constants that are the input to our results, we refer the reader to, e.g., [1].

In addition to concentration inequalities, we also prove a robustness result for the dynamics, with respect to model-form uncertainty. For such uncertainty quantification (UQ) bounds, we think of (X_t, P^μ), called the baseline model, as an imperfect representation of a "true" (or, at least, more precise) alternative model. This alternative model may not be fully known, or it might be intractable (analytically or numerically), and so one may want to investigate how sensitive the results for the baseline model are to (not necessarily small) model perturbations. The next theorem provides such performance guarantees, generalizing the results in [7], and is based on the general approach to uncertainty quantification introduced in [12] and further developed in [6, 20, 29, 35]. In this

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

context, the goal is to control the bias

$$\widetilde{E}^{\widetilde{\mu}}\left[\frac{1}{T}\int_0^T f(\widetilde{X}_t)\,dt\right] - \mu_*[f],$$

where $(\widetilde{X}_t, \widetilde{P}^{\widetilde{\mu}})$ with $\widetilde{X}_0 \sim \widetilde{\mu}$ is the alternative model and $\widetilde{E}^{\widetilde{\mu}}$ is the expectation with respect to $\widetilde{P}^{\widetilde{\mu}}$. We denote by P_T^{μ} and $\widetilde{P}_T^{\widetilde{\mu}}$ the path-space distributions of the base and alternative models on the time window $[0, T]$ and prove the following result in Section 3 (see Theorem 3.5).

THEOREM 1.2 (Uncertainty quantification bounds). Suppose that the baseline Markov process X_t satisfies the hypocoercive estimate (1.4) and $(\widetilde{X}_t, \widetilde{P}^{\widetilde{\mu}})$ is a stochastic process such that the path space relative entropy satisfies

$$R\big(\widetilde{P}_T^{\widetilde{\mu}} \| P_T^{\mu_*}\big) < \infty. \tag{1.10}$$

Then for any bounded observable f and any time $T > 0$, we have

$$\left|\widetilde{E}^{\widetilde{\mu}}\left[\frac{1}{T}\int_0^T f(\widetilde{X}_t)\,dt\right] - \mu_*[f]\right| \leq \sqrt{2v\eta_T} + b\eta_T, \tag{1.11}$$

where v and b are given in (1.9), and

$$\eta_T = \frac{1}{T}\left(\log\left((1-\epsilon)^{-1/2}\right) + R\big(\widetilde{P}_T^{\widetilde{\mu}} \| P_T^{\mu_*}\big)\right).$$

Note that if the perturbed dynamics $(\widetilde{X}_t, \widetilde{P}^{\widetilde{\mu}})$ is itself ergodic (Markovian or not), one can often prove that the entropy production rate $\lim_{T\to\infty} \frac{1}{T} R\big(\widetilde{P}_T^{\widetilde{\mu}} \| P_T^{\mu_*}\big)$ exists and is finite. In that case, Theorem 1.2 provides an uncertainty quantification bound for the expectation of f under corresponding stationary distributions.

The remainder of this paper is organized as follows. In Section 2 we review several examples of hypocoercive systems to which our results apply. There, we also give an overview of the hypocoercivity method of [15, 16]. This method is a crucial tool in the proofs of our new results, namely the concentration inequalities and UQ bounds outlined above; proofs of these are given in Section 3.

2. Hypocoercive MCMC samplers. In this section we introduce several examples of popular hypocoercive samplers for which the modified Poincaré inequality (1.4) has been proven by following the strategy of [15, 16]. In particular we consider several examples of partially deterministic MCMC samplers studied in [1]. We will refer the reader to the original papers for technical details and content ourselves with a brief, and at times somewhat informal, overview:

Consider a probability measure $d\nu_*(q) = Z^{-1}e^{-\beta V(q)}dq$ on $\mathbb{R}^d$ to be sampled and for which a Poincaré inequality holds; i.e., there exists a constant $C_{\nu_*} > 0$ such that for all $g \in L^2(\nu_*)$ we have

$$\|\nabla_q g\|^2_{L^2(\nu_*)} \geq C_{\nu_*} \operatorname{Var}_{\nu_*}[g]. \tag{2.1}$$

See, e.g., [2] for conditions on V which imply a Poincaré inequality.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

Define the product measure $\mu_* = \nu_* \times \rho_*$ on the extended phase space $\mathbb{R}^d \times \mathcal{P}$ and the projection $\Pi f = \int f d\rho_*$. We consider a Markov process $X_t = (Q_t, P_t)$ on $\mathbb{R}^d \times \mathcal{P}$ with invariant measure μ_* and assume standard smoothness and growth conditions on V to ensure that X_t induces a strongly continuous semigroup T_t on $L^2(\mu_*)$ with generator A, and with the time-reversed process having generator given by the adjoint A^* of A on $L^2(\mu_*)$. We decompose A into symmetric and antisymmetric parts:

$$A = S + T, \quad \text{with } S = \frac{A + A^*}{2} \quad \text{and} \quad T = \frac{A - A^*}{2}. \tag{2.2}$$

The following four examples fit within this framework and that will be used to illustrate the utility of our results; see [1] for a proof of hypocoercivity of a more general class of models which covers all examples considered here, as well as [34, 45, 46] for further examples (some of them being nonequilibrium as well).

(1) **(Langevin and modified Langevin equations)** The (underdamped) Langevin equation is the system of stochastic differential equations on $\mathbb{R}^{2d}$ given by

$$dQ_t = \frac{P_t}{m} dt, \quad dP_t = \left(-\nabla V(Q_t) - \gamma \frac{P_t}{m} \right) dt + \sqrt{\frac{2\gamma}{\beta}} dW_t, \tag{2.3}$$

where $m > 0$ is the mass, $\beta > 0$ is proportional to the inverse temperature, $\gamma > 0$ is the drag coefficient, W_t is a Wiener process, and $V : \mathbb{R}^d \to \mathbb{R}$ is a smooth potential. The appropriate ρ_* is a Gaussian measure with mean 0 and covariance matrix $m/\beta I$. The generator A is an extension of the differential operator

$$A = \underbrace{\frac{\gamma}{\beta} \Delta_p - \gamma \left(\frac{p}{m} \right)^T \nabla_p}_{=S} + \underbrace{\left(\frac{p}{m} \right)^T \nabla_q - \nabla V(q)^T \nabla_p}_{=T}. \tag{2.4}$$

This is the model originally considered in [15, 16] and several modifications of this models have also been shown to be hypocoercive. For example [41, 45] consider a Langevin equation with a modified kinetic energy (nonquadratic) so that ρ_* is not Gaussian and the diffusion does not need to be hypoelliptic. Further generalizations of the Langevin equations with general ρ_* are also considered in [1].

(2) **(Hybrid Hamiltonian Monte Carlo)** In this randomized version of Hamiltonian Monte-Carlo introduced by [17], the system follows Hamiltonian equations of motion with Hamiltonian $V(q) + p^2/2m$ for an exponentially distributed amount of time, after which the momentum is resampled from the Gaussian measure ρ_*. The generator has the form

$$A = \underbrace{\lambda(\Pi - I)}_{=S} + \underbrace{\left(\frac{p}{m} \right)^T \nabla_q - \nabla V^T \nabla_p}_{=T}. \tag{2.5}$$

(3) **(Bouncy particle sampler)** In this sampler, introduced originally in [40], a particle starting at time t_0 in the state (q_0, p_0) moves freely $p(t) = p(t_0)$ and $q(t) = q(t_0) + t\frac{p(t_0)}{m}$ up to the random time $t_0 + \tau$. The updating time τ is governed by two mechanisms: either the velocity of the particle is refreshed, i.e., p

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

is sampled from the Gaussian ρ_* (this occurs at rate λ), or the particle "bounces", i.e., it undergoes a Newtonian elastic collision on the hyperplane tangential to the gradient of the energy and the momentum is updated according to the rule

$$R(q)p = p - \frac{p^T \nabla V(q)}{\|\nabla V\|^2} \nabla V. \tag{2.6}$$

The time at which this happens is governed by an inhomogeneous Poisson process of intensity $\lambda(q,p) = \left[\left(\frac{p}{m}\right)^T \nabla V(q)\right]^+$. If we set $Rf(q,p) = f(q, R(q)p)$, then the generator is

$$A = \left(\frac{p}{m}\right)^T \nabla_q + \left[\left(\frac{p}{m}\right)^T \nabla V(q)\right]^+ (R - I) + \lambda(\Pi - I), \tag{2.7}$$

and elementary computations shows that $\mu_* = \nu_* \times \rho_*$ is invariant and

$$S = \left|\left(\frac{p}{m}\right)^T \nabla V(q)\right| (R - I) + \lambda(\Pi - I), \tag{2.8}$$

$$T = \left(\frac{p}{m}\right)^T \nabla_q + \left(\frac{p}{m}\right)^T \nabla V(q)(R - I). \tag{2.9}$$

(4) **(Zig-Zag Sampler)** In the zig-zag sampler, contrary to the other examples, the velocity is discrete, and, for example, ρ_* is the uniform distribution on $\{-1, 1\}^d$. As in the bouncy sampler, the trajectories are piecewise linear. At updating times, the (randomly chosen) ith component of the velocity is reversed; see [4] for a more detailed discussion. The generator of the Markov process has the form

$$A = v^T \nabla_q + \sum_{i=1}^{d} \left[v_i \partial_{q_i} V(q)\right]^+ (R_i - I) + \lambda(\Pi - I), \tag{2.10}$$

where $R_i f(q,v) = f(q, v - 2(e_i^T v)e_i)$ (with e_i the standard basis vector in $\mathbb{R}^d$). A computation similar to the one for the bouncy sampler shows that

$$S = \sum_{i=1}^{d} \left|v_i \partial_{q_i} V(q)\right| (R_i - I) + \lambda(\Pi - I), \tag{2.11}$$

$$T = v \nabla_q + \sum_{i=1}^{d} v_i \partial_{q_i} V(q)(R_i - I). \tag{2.12}$$

Note that for all the examples considered, it is easy to verify that one has the identity

$$T\Pi = \frac{p}{m} \nabla_q \Pi \tag{2.13}$$

(with the convention that $p/m = v$ for the zig-zag sampler). This fact is used to establish the following functional analytic estimates (see [15, 16]) which are the basis for the hypocoercive estimates (for the convenience of the reader the proof is in Appendix A).

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

PROPOSITION 2.1. Define

$$B = (I + (T\Pi)^*(T\Pi))^{-1}(-T\Pi)^*. \tag{2.14}$$

The operators S, T, and B have the following properties:

(1) $B1 = B^*1 = 0$,
(2) $S = (I - \Pi)S(I - \Pi)$,
(3) $T\Pi = (I - \Pi)T\Pi$,
(4) $B = \Pi B = \Pi B(I - \Pi)$, and B and TB are bounded operators with $\|Bf\| \leq 1/2\|(I - \Pi)f\|$ and $\|TBf\| \leq \|(I - \Pi)f\|$.

Next, define the family of modified scalar products on $L^2(\mu_*)$,

$$\langle f, g\rangle_\epsilon = \langle f, g\rangle + \epsilon\langle f, (B + B^*)g\rangle, \ \ \epsilon \in (0, 1). \tag{2.15}$$

As $\|B\| \leq 1/2$, $\langle\cdot,\cdot\rangle_\epsilon$ is an inner product whose induced norm is equivalent to that of $\langle\cdot,\cdot\rangle$. As a consequence of Proposition 2.1 one obtains for suitable f that satisfy $\mu_*[f] = 0$

$$\begin{aligned}\langle Af, f\rangle_\epsilon =& \langle Sf, f\rangle + \epsilon\left[\langle BSf, f\rangle + \langle BTf, f\rangle + \langle SBf, f\rangle - \langle TBf, f\rangle\right] \\ =& \langle (S - \epsilon TB)(I - \Pi)f, (I - \Pi)f\rangle + \epsilon\langle BT\Pi f, \Pi f\rangle \\ &+ \epsilon\left[\langle BS(I - \Pi)f, \Pi f\rangle + \langle BT(I - \Pi)f, \Pi f\rangle\right],\end{aligned} \tag{2.16}$$

where we have used that $SB = 0$. The various terms in (2.16) can be bounded as follows:

(1) The term $\langle (S - \epsilon TB)(I - \Pi)f, (I - \Pi)f\rangle$ is controlled by the dissipative term in the p-variables (since TB is bounded) and it is not difficult to see that in the cases considered here we have a Poincaré inequality in the p-variables (averaged over ν_*):

$$\langle f, -Sf\rangle \geq \lambda_p\|(I - \Pi)f\|^2 \tag{2.17}$$

for some $\lambda_p > 0$. For the Langevin equation $\lambda_p = \frac{\gamma}{\beta}$ is the spectral gap of the Ornstein-Uhlenbeck process, while for the other examples we can take $\lambda_p = \lambda$ from the velocity resampling mechanism.

(2) For the term $\langle BT\Pi f, \Pi f\rangle$, note that using (2.13) together with the Poincaré inequality for the measure ν_*, we have

$$\langle f, (T\Pi^*T\Pi)f\rangle = \Pi\left(\frac{p^2}{m^2}\right)\|\nabla_q\Pi f\|^2 \geq \Pi\left(\frac{p^2}{m^2}\right)C_{\nu_*}\|\Pi f\|^2, \tag{2.18}$$

where C_{ν_*} is the Poincaré constant for the measure ν_*. Then, since $-BT\Pi = (I + (T\Pi)^*T\Pi)^{-1}(T\Pi)^*T\Pi$, by functional calculus we have

$$\langle -BT\Pi f, \Pi f\rangle \geq \left(1 - \left(1 + \Pi\left(\frac{p^2}{m^2}\right)C_{\nu_*}\right)^{-1}\right)\|\Pi f\|^2 \equiv \lambda_q\|\Pi f\|^2, \tag{2.19}$$

where $\lambda_q \in (0, 1)$.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

(3) For the off-diagonal terms it is enough to show that they are bounded; i.e.,

$$\|BT(I-\Pi)f\| + \|BS(I-\Pi)f\| \le R_0\|(I-\Pi)f\|. \tag{2.20}$$

The bound of the first term is the technical part of the proof. For the Langevin equation this is proved in [16], for the other samplers it is proved in [1] (see Lemma 29 and Lemma 32 in [1] in particular and the bound in Lemma 11 in [1] which is specific to the zig-zag sampler).

Based on these estimates, one has constants $\lambda_q, \lambda_p, R_0 > 0$ such that for any f with $\mu_*[f] = 0$,

$$\begin{aligned}\langle -Af, f\rangle_\epsilon &\ge \begin{bmatrix} \|\Pi f\| \\ \|(I-\Pi)f\| \end{bmatrix}^T \begin{bmatrix} \epsilon\lambda_q & -\epsilon R_0/2 \\ -\epsilon R_0/2 & \lambda_p - \epsilon \end{bmatrix} \begin{bmatrix} \|\Pi f\| \\ \|(I-\Pi)f\| \end{bmatrix} \\ &\ge \Lambda(\epsilon)\,\mathrm{Var}_{\mu_*}[f],\end{aligned} \tag{2.21}$$

where

$$\Lambda(\epsilon) \equiv \frac{(\lambda_q - 1)\epsilon + \lambda_p - \sqrt{((\lambda_q+1)\epsilon - \lambda_p)^2 + \epsilon^2 R_0^2}}{2} \tag{2.22}$$

is the smallest eigenvalue of the matrix in equation (2.21). Note that $\Lambda(0) = 0$ and $\Lambda'(0) = \lambda_q > 0$; therefore, $\Lambda(\epsilon) > 0$ for ϵ sufficiently small. Specifically, $\Lambda(\epsilon)$ is positive if

$$0 < \epsilon < 4\lambda_q\lambda_p/(4\lambda_q + R_0^2). \tag{2.23}$$

In the next section, we show how the Poincaré inequality (2.21) for the modified inner product (2.15) can be used to derive nonasymptotic confidence intervals and UQ bounds for hypocoercive systems, having in mind the four examples outlined above.

3. Concentration inequalities and UQ bounds via Feynman-Kac semigroups. In this section, we prove our main new results for hypocoercive systems:

(1) A concentration inequality and corresponding nonasymptotic confidence intervals in Section 3.2.
(2) UQ bounds in Section 3.3.

The former are obtained by an adaptation of the technique from [49] and [28] to hypocoercive systems, which we first summarize.

3.1. *Background.* As in [28, 49], we will prove Bernstein-type concentration inequalities. The following related elementary facts will be used repeatedly (see, e.g., the discussion of subgamma random variables in Chapter 2 in [10]):

Consider the convex function $\Psi_{v,b}$ given by

$$\Psi_{v,b}(\lambda) = \frac{\lambda^2 v}{2(1-\lambda b)} \quad \text{for } 0 \le \lambda < 1/b. \tag{3.1}$$

Its (one-sided) Legendre transform $\Psi^*_{v,b}$ is

$$\Psi^*_{v,b}(r) = \sup_{0\le\lambda<1/b}\{\lambda r - \Psi_{v,b}(\lambda)\} = \frac{2r^2}{v\left(1 + \sqrt{1 + \frac{2br}{v}}\right)^2} \quad \text{for } r \ge 0 \tag{3.2}$$

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

and the inverse of the Legendre transform $\Psi^*_{v,b}$ is

$$(\Psi^*_{v,b})^{-1}(\eta) = \inf_{\lambda>0}\left\{\frac{\Psi_{v,b}(\lambda)+\eta}{\lambda}\right\} = \sqrt{2v\eta}+b\eta \quad \text{for } \eta \geq 0. \tag{3.3}$$

Now we summarize the method of [28, 49]:
Let $\mathcal{X}$ be a Polish space and suppose we have time homogeneous, $\mathcal{X}$-valued, càdlàg Markov processes $(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, P^x)$, $x \in \mathcal{X}$, with initial distributions $(X_0)_* P^x = \delta_x$ for all x. For an initial measure μ, write $P^\mu = \int P^x d\mu(x)$.

We assume that μ_* is an invariant ergodic measure on $\mathcal{X}$ and we consider the real Hilbert space $L^2(\mu_*)$ with scalar product $\langle \cdot, \cdot \rangle$. Define the strongly continuous Markov semigroup $T_t : L^2(\mu_*) \to L^2(\mu_*)$ by

$$T_t[f](x) = E^x[f(X_t)] \tag{3.4}$$

and denote its generator by $(A, D(A))$.

More generally, for a bounded measurable $V : \mathcal{X} \to \mathbb{R}$, define the Feynman-Kac semigroup $T^V_t : L^2(\mu_*) \to L^2(\mu_*)$ by

$$T^V_t[f](x) = E^x\left[f(X_t)e^{\int_0^t V(X_s)ds}\right], \tag{3.5}$$

which is a strongly continuous semigroup with generator $(A+V, D(A))$. If we set

$$\kappa(V) \equiv \sup\left\{\langle (A+V)g, g\rangle : g \in D(A), \|g\| = 1\right\}, \tag{3.6}$$

then, by definition (and as long as $\kappa(V) < \infty$), for any $g \in D(A)$ we have

$$\langle (A + V - \kappa(V))g,\, g\rangle \leq 0 \tag{3.7}$$

and thus by the Lumer-Philipps theorem (see, e.g., Chapter IX in [50]) the semigroup generated by $A + V - \kappa(V)$ is a contraction semigroup on $L^2(\mu_*)$. This implies that

$$\|T^V_t\| \leq e^{t\kappa(V)}, \;\; t \geq 0 \tag{3.8}$$

(note that equation (3.8) also trivially holds if $\kappa(V) = \infty$). Therefore, by the Chernoff bound we have

$$\begin{aligned} P^\mu\left(\frac{1}{T}\int_0^T f(X_t)dt - \mu_*[f] > r\right) &\leq \inf_{\lambda>0} e^{-\lambda T r} E^\mu\left[e^{\lambda\int_0^T \widehat{f}(X_t)dt}\right] \\ &\leq \inf_{\lambda>0} e^{-\lambda T r}\int T^{\lambda\widehat{f}}_T[1]d\mu \\ &\leq \inf_{\lambda>0} e^{-\lambda T r}\left\|\frac{d\mu}{d\mu_*}\right\|\left\|T^{\lambda\widehat{f}}_T\right\| \\ &\leq \left\|\frac{d\mu}{d\mu_*}\right\| e^{-T\sup_{\lambda>0}\{\lambda r - \kappa(\lambda\widehat{f})\}}. \end{aligned} \tag{3.9}$$

This basic insight, first noted in [49], can also be extended to unbounded V. From here, one can obtain explicit concentration inequalities by further bounding $\kappa(\lambda\widehat{f})$ (which contains the Dirichlet form $\langle Ag, g\rangle$) using $L^2(\mu_*)$-functional inequalities, such as a Poincaré inequality (or log-Sobolev inequalities, Lyapunov functions, and so on); see [11, 28, 30, 36, 49] for many such examples.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

3.2. *Concentration inequalities.* In the hypocoercive examples considered in this paper, the generator is nonreversible and there is no Poincaré inequality with respect to the $L^2(\mu_*)$-scalar product but, as discussed in Section 2, there is a Poincaré inequality in terms of a modified scalar product that induces an equivalent norm. In the following theorem, we show that one still obtains concentration inequalities in this more general setting.

THEOREM 3.1 (Concentration inequalities). Let $(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, P^x)$, $x \in \mathcal{X}$, be $\mathcal{X}$-valued càdlàg Markov processes with invariant ergodic measure μ_*.

Let $\langle \cdot, \cdot \rangle_\#$ be an inner product on $L^2(\mu_*)$ such that:

(1) The induced norms $\| \cdot \|_\#$ and $\| \cdot \|$ are equivalent: there exists $0 < c \leq C < \infty$ such that $c\| \cdot \| \leq \| \cdot \|_\# \leq C\| \cdot \|$.
(2) For all $g \in L^2(\mu_*)$, we have $\langle g, 1 \rangle_\# = \langle g, 1 \rangle$.
(3) A Poincaré inequality holds for $\langle \cdot, \cdot \rangle_\#$; i.e., we have $\alpha > 0$ such that

$$\|g\|_\#^2 \leq \alpha \langle -Ag, g \rangle_\# \quad \text{for all } g \in D(A) \text{ with } \mu_*[g] = 0. \tag{3.10}$$

For bounded measurable f, let $M_{\widehat{f}}$ denote the multiplication operator by $\widehat{f} = f - \mu_*[f]$. We have the following concentration inequalities for $T > 0$:

$$P^\mu \left(\pm \left[\frac{1}{T} \int_0^T f(X_t) dt - \mu_*[f] \right] \geq r \right) \leq c^{-1} \left\| \frac{d\mu}{d\mu_*} \right\| e^{-T\Psi^*_{v_\pm, b_\pm}(r)}, \tag{3.11}$$

where $\Psi^*_{\nu,b}$ is given in (3.2),

$$v_\pm = 2\alpha \left\| \frac{1}{2}(M_{\pm\widehat{f}} + M^\dagger_{\pm\widehat{f}})1 \right\|_\#^2, \qquad b_\pm = \alpha \max \left\{ 0, \sup_{\|g\|_\# = 1} \langle M_{\pm\widehat{f}} g, g \rangle_\# \right\}, \tag{3.12}$$

and $M^\dagger_{\widehat{f}}$ denotes the adjoint with respect to the $\langle \cdot, \cdot \rangle_\#$-inner product.

Note that $\nu_\pm$ and $b_\pm$ can be replaced by any upper bounds on these quantities, for example in terms of the $L^2(\mu_*)$-norm (see the calculation for the hypocoercive examples in Section 3.4).

Proof. The proof is a modification of the strategy used in [28]. We start as in equation (3.9) but use the Lumer-Phillips theorem for the $\| \cdot \|_\#$ norm instead since, by equivalence of the norms, $T_t^{\lambda \widehat{f}}$ is also a strongly continuous semigroup on $(L^2(\mu_*), \| \cdot \|_\#)$ with the same generator. Using the Chernoff bound, the equivalence of the norm, and the fact

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

that, by assumption (2), $\|1\|_\# = \langle 1,1\rangle_\# = \langle 1,1\rangle = 1$, we obtain

$$\begin{aligned} P^\mu\left(\frac{1}{T}\int_0^T f(X_t)dt \geq \mu_*[f] + r\right) &\leq \inf_{\lambda>0} e^{-\lambda T r} E^\mu\left[e^{\lambda\int_0^T \widehat{f}(X_t)dt}\right] \\ &= \inf_{\lambda>0} e^{-\lambda T r}\int T_T^{\lambda\widehat{f}}[1]\frac{d\mu}{d\mu_*}d\mu_* \\ &\leq \inf_{\lambda>0} e^{-\lambda T r}\left\|\frac{d\mu}{d\mu_*}\right\|\left\|T_T^{\lambda\widehat{f}}[1]\right\| \\ &\leq \inf_{\lambda>0} e^{-\lambda T r}\left\|\frac{d\mu}{d\mu_*}\right\| c^{-1}\left\|T_T^{\lambda\widehat{f}}\right\|_\# \|1\|_\# \\ &\leq c^{-1}\left\|\frac{d\mu}{d\mu_*}\right\| e^{-T\sup_{\lambda>0}(\lambda r - \kappa_\#(\lambda\widehat{f})}, \end{aligned} \tag{3.13}$$

where, by the Lumer-Phillips theorem applied to $L^2(\mu_*)$ with the scalar product $\langle\cdot,\cdot\rangle_\#$,

$$\kappa_\#(\lambda\widehat{f}) \equiv \sup\left\{\langle (A+\lambda\widehat{f})g, g\rangle_\# : g\in D(A), \|g\|_\# = 1\right\}. \tag{3.14}$$

Next we use Lemma 3.2 proved in [7], which is a generalization of a result in [28], which itself was a simplification of the argument originally used in [36]. For completeness, the proof is given in Appendix A.

LEMMA 3.2. Let H be a real Hilbert space, let $A : D(A)\subset H\to H$ be a linear operator, and let $M : H\to H$ be a bounded linear operator; denote its adjoint by $M^\dagger$. Assume there exist $\alpha>0$ and $x_0\in H$ with $\|x_0\| = 1$ such that

$$\langle Mx_0, x_0\rangle = 0 \quad\text{and}\quad \langle Ax, x\rangle \leq -\alpha^{-1}\|P^\perp x\|^2 \tag{3.15}$$

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

$$\sup_{x\in D(A),\|x\|=1}\langle (A+\lambda M)x, x\rangle \leq \frac{\lambda^2\alpha V}{1-\lambda\alpha K} = \Psi_{2\alpha V,\alpha K}(\lambda) \tag{3.16}$$

for $0\leq\lambda<1/\alpha K$, where

$$V = \left\|\frac{1}{2}(M+M^\dagger)x_0\right\|^2, \quad K = \max\left\{0, \sup_{\|y\|=1}\langle My, y\rangle\right\}. \tag{3.17}$$

To use this result we take $H = (L^2(\mu_*), \langle\cdot,\cdot\rangle_\#)$ and let A be the generator, $M = M_{\widehat{f}}$, and $x_0 = 1$. By assumption (2), we have $\langle Mx_0, x_0\rangle_\# = \langle\widehat{f}, 1\rangle_\# = \langle\widehat{f}, 1\rangle = 0$. This assumption also implies that the projection onto $1^\perp$ (for both scalar products) is given by $P^\perp f = \widehat{f}$ and

$$\langle Ag, 1\rangle_\# = \langle Ag, 1\rangle = 0, \quad g\in D(A). \tag{3.18}$$

Combined with assumption (3) and the fact that $A[1] = 0$, we get

$$\langle Ag, g\rangle_\# = \langle A\widehat{g}, \widehat{g}\rangle_\# \leq -\alpha^{-1}\|\widehat{g}\|_\#^2 = -\alpha^{-1}\|P^\perp g\|_\#^2,$$

and thus we can apply Lemma 3.2 to obtain

$$\kappa_\#(\lambda\widehat{f}) = \sup_{g\in D(A),\|g\|_\#=1}\langle (A+\lambda\widehat{f})g, g\rangle_\# \leq \Psi_{v_+,b_+}(\lambda) \tag{3.19}$$

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

for all $0 \le \lambda < 1/b_+$, where

$$v_+ = 2\alpha \left\| \frac{1}{2}(M_{\widehat{f}} + M_{\widehat{f}}^\dagger)1 \right\|_\#^2, \qquad b_+ = \alpha \max\left\{0, \sup_{\|g\|_\#=1} \langle M_{\widehat{f}} g, g\rangle_\#\right\} \tag{3.20}$$

(as was given in (3.12)). Therefore,

$$\begin{aligned} P^\mu\left(\frac{1}{T}\int_0^T f(X_t)dt \ge \mu_*[f] + r\right) &\le c^{-1}\left\|\frac{d\mu}{d\mu_*}\right\| e^{-T\sup_{0\le\lambda<1/b_+}\{\lambda r - \Psi_{v_+,b_+}(\lambda)\}} \\ &= c^{-1}\left\|\frac{d\mu}{d\mu_*}\right\| e^{-T\Psi^*_{v_+,b_+}(r)}. \end{aligned} \tag{3.21}$$

The lower bound is obtained by replacing f by $-f$ and this concludes the proof. □

As an immediate corollary we obtain a nonasymptotic confidence interval.

COROLLARY 3.3 (Confidence intervals). Under the same assumptions as in Theorem 3.1, given a time T and a confidence level $0 < 1 - \delta < 1$ we have

$$P^\mu\left(\frac{1}{T}\int_0^T f(X_t)dt - \mu_*[f] \in (-r_-, r_+)\right) \ge 1 - \delta \tag{3.22}$$

where

$$r_\pm = \sqrt{2v_\pm \frac{1}{T}\log\left(\frac{2N}{\delta}\right)} + b_\pm \frac{1}{T}\log\left(\frac{2N}{\delta}\right), \tag{3.23}$$

with $N = c^{-1}\left\|\frac{d\mu}{d\mu_*}\right\|$ and $v_\pm$ and $b_\pm$ given in (3.12).

Proof. Define $\eta = \frac{1}{T}\log\left(\frac{2N}{\delta}\right)$ (note that $N \ge 1$ follows from assumptions (1) and (2) of Theorem 3.1), so that $r_\pm = (\Psi^*_{v_\pm,b_\pm})^{-1}(\eta)$, with $r_\pm$ given as in (3.23). Using $r = r_\pm$ in the concentration bound in Theorem 3.1 we find

$$P^\mu\left(\pm\left[\frac{1}{T}\int_0^T f(X_t)dt - \mu_*[f]\right] \ge r_\pm\right) \le Ne^{-T\eta} = \frac{\delta}{2}. \tag{3.24}$$

The result (3.22) then follows from a union bound. □

3.3. *Robustness bounds on steady state bias due to model-form uncertainty.* Following the methods in [7, 29], we can also use the above tools to obtain bounds on the bias of the expectation of ergodic averages when the process itself is subject to (model-form) uncertainty.

We think of the Markov process (X_t, P^μ) considered in Section 3.1 as the baseline process and consider an alternative stochastic process $(\widetilde{X}_t, \widetilde{P}^{\widetilde{\mu}})$ with initial distribution $(X_0)_*\widetilde{P}^{\widetilde{\mu}} = \widetilde{\mu}$ and let $\widetilde{E}^{\widetilde{\mu}}$ be the associated expectation.

REMARK 3.4. The requirements on the alternative process are very minimal. In particular, we are *not* assuming $(\widetilde{X}_t, \widetilde{P}^{\widetilde{\mu}})$ is a Markov process.

We will compare the two processes using relative entropy; we assume absolute continuity of the path-space distributions on finite time windows $[0, T]$; i.e., $\widetilde{P}^{\widetilde{\mu}}_T \ll P^\mu_T$, and also assume the relative entropy is finite:

$$R\big(\widetilde{P}^{\widetilde{\mu}}_T \| P^\mu_T\big) < \infty. \tag{3.25}$$

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

See the supplementary material to [20] for a collection of techniques that can be used to bound the path-space relative entropy (3.25) for various classes of alternative models.

Given an observable f, we consider the ergodic averages

$$\widetilde{F}_T = \frac{1}{T}\int_0^T f(\widetilde{X}_t)dt, \quad F_T = \frac{1}{T}\int_0^T f(X_t)dt, \tag{3.26}$$

and are interested in bounding the bias between the baseline and the alternative processes:

$$\widetilde{E}^{\widetilde{\mu}}[\widetilde{F}_T] - E^{\mu}[F_T]. \tag{3.27}$$

THEOREM 3.5 (Uncertainty quantification bounds). Let (X_t, P^x), $x \in \mathcal{X}$, be a family of Markov processes satisfying the assumptions of Theorem 3.1, let μ be an initial distribution, and let $(X_t, \widetilde{P}^{\widetilde{\mu}})$ be an alternative process with $R(\widetilde{P}_T^{\widetilde{\mu}} \| P_T^{\mu}) < \infty$. Then for any bounded measurable f we have

$$\pm\left(\widetilde{E}^{\widetilde{\mu}}[\widetilde{F}_T] - E^{\mu}[F_T]\right) \leq \sqrt{2v_{\pm}\eta_T} + b_{\pm}\eta_T + \frac{C}{c}\frac{1-e^{-\alpha T}}{T}\left\|\frac{d\mu}{d\mu_*}\right\| \mathrm{Var}_{\mu_*}[f],$$

where $v_{\pm}$ and $b_{\pm}$ are given in (3.12) and

$$\eta_T = \frac{1}{T}\left(\log(c^{-1}) + \log\left\|\frac{d\mu}{d\mu_*}\right\| + R\big(\widetilde{P}_T^{\widetilde{\mu}} \| P_T^{\mu}\big)\right).$$

If, in addition, the process $(\widetilde{X}_t, \widetilde{P}^{\widetilde{\mu}})$ is ergodic with invariant measure $\widetilde{\mu}_*$, the limit

$$\eta_\infty = \lim_{T\to\infty} \frac{1}{T} R\big(\widetilde{P}_T^{\widetilde{\mu}} \| P_T^{\mu_*}\big) \tag{3.28}$$

exists for the relative entropy rate, and $\|d\mu/d\mu_*\| < \infty$, then we have the steady state bias bound

$$\pm\left(\widetilde{\mu}_*[f] - \mu_*[f]\right) \leq \sqrt{2v_{\pm}\eta_\infty} + b_{\pm}\eta_\infty. \tag{3.29}$$

Proof. The proof proceeds along the same line as in [7] to which we refer for more details. The starting point is the Gibbs information inequality [12, 20]: for g bounded and measurable and probability measures Q and $\widetilde{Q}$

$$\pm\left(E_{\widetilde{Q}}E[g] - E_Q[g]\right) \leq \inf_{\lambda>0}\left\{\frac{\log E_Q[e^{\pm\lambda(g-E_Q[g])}] + R(\widetilde{Q}\|Q)}{\lambda}\right\}. \tag{3.30}$$

This is a direct consequence of the Gibbs variational principle for the relative entropy [19].

We apply the bound to the measures P_T^{μ}, $\widetilde{P}_T^{\mu}$ (distributions on path-space up to time T) and $g(x) = \int_0^T f(x_t)dt$ (a bounded measurable function of paths, x, up to time T)

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

and then divide both sides by T:

$$\begin{aligned}
&\pm\left(\widetilde{E}^{\widetilde{\mu}}[\widetilde{F}_T] - E^{\mu}[F_T]\right) \\
&\le \inf_{\lambda>0}\left\{\frac{\log E^{\mu}[e^{\pm\lambda T(F_t-\mu_*[f])-E^{\mu}[F_T]}] + R(\widetilde{P}_T^{\widetilde{\mu}}\|P_T^{\mu})}{\lambda T}\right\} \\
&\le \underbrace{\inf_{\lambda>0}\left\{\frac{\log E^{\mu}[e^{\pm\lambda T(F_t-\mu_*[f])}] + R(\widetilde{P}_T^{\widetilde{\mu}}\|P_T^{\mu})}{\lambda T}\right\}}_{=(\mathrm{I})} \mp \underbrace{(E^{\mu}[F_T]-\mu_*[f])}_{=(\mathrm{II})}.
\end{aligned} \tag{3.31}$$

The term (II) only involves the baseline process and is easily bounded, for example using the Poincaré inequality for the scalar product $\langle\cdot,\cdot\rangle_\epsilon$:

$$\begin{aligned}
|(\mathrm{II})| = \left|E^{\mu}\left[\frac{1}{T}\int_0^T \widehat{f}(X_t)\,dt\right]\right| &\le \frac{1}{T}\int_0^T \left|\left\langle \frac{d\mu}{d\mu_*}, T_t[\widehat{f}]\right\rangle\right| dt \\
&\le \frac{1}{T}\int_0^T e^{-t/\alpha}\left\|\frac{d\mu}{d\mu_*}\right\|\frac{C}{c}\|\widehat{f}\|dt \\
&= \frac{C}{c}\frac{1-e^{-T/\alpha}}{T/\alpha}\left\|\frac{d\mu}{d\mu_*}\right\|\sqrt{\mathrm{Var}_{\mu_*}[f]}.
\end{aligned} \tag{3.32}$$

To bound the term (I), we use Lemma 3.2 to bound the moment generating function, similarly to the proof of Theorem 3.1:

$$\begin{aligned}
(\mathrm{I}) &= \inf_{\lambda>0}\left\{\frac{\log\int T_T^{\pm\lambda\widehat{f}}[1]d\mu + R(\widetilde{P}_T^{\widetilde{\mu}}\|P_T^{\mu})}{\lambda T}\right\} \\
&\le \inf_{\lambda>0}\left\{\frac{\log\left(c^{-1}\left\|\frac{d\mu}{d\mu_*}\right\| e^{T\kappa_{\#}(\pm\lambda\widehat{f})}\right) + R(\widetilde{P}_T^{\widetilde{\mu}}\|P_T^{\mu})}{\lambda T}\right\} \\
&= \inf_{\lambda>0}\left\{\frac{\kappa_{\#}(\pm\lambda\widehat{f})+\eta_T}{\lambda}\right\} \\
&\le \inf_{\lambda>0}\left\{\frac{\Psi_{v_\pm,b_\pm}(\lambda)+\eta_T}{\lambda}\right\} \\
&= (\Psi^*_{v_\pm,b_\pm})^{-1}(\eta_T) = \sqrt{2v_\pm\eta_T} + b_\pm\eta_T.
\end{aligned} \tag{3.33}$$

Finally, by taking $T\to\infty$ we obtain the bounds in equation (3.29) □

3.4. *Application to hypocoercive samplers.* Theorems 1.1 and 1.2 for hypocoercive MCMC samplers follow rather immediately from Corollary 3.3 and from Theorem 3.5. We first verify the three assumptions in Theorem 3.1. The modified scalar product (2.15) has the form $\langle f,g\rangle_\epsilon = \langle f,g\rangle + \epsilon\langle f, Gg\rangle$ where $G1=0$ and $\|G\|\le 1$. Therefore, we have $c=(1-\epsilon)^{1/2}$, $C=(1+\epsilon)^{1/2}$, and $\langle f,1\rangle_\epsilon = \langle f,1\rangle$, and, for $\epsilon\in(0,1)$ sufficiently small (see equation (2.23)), by equation (2.21) we have $\alpha = \frac{1+\epsilon}{\Lambda(\epsilon)}$.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

Since $\langle M_{\widehat{f}} g, g\rangle_\epsilon \le \|\widehat{f}\|_\infty \|g\|^2(1+\epsilon) \le \frac{1+\epsilon}{1-\epsilon}\|\widehat{f}\|_\infty \|g\|_\epsilon^2$ we have

$$b_\pm = \alpha \max\left\{0, \sup_{\|g\|_\epsilon=1} \langle M_{\pm\widehat{f}} g, g\rangle_\epsilon\right\} \le \frac{(1+\epsilon)^2}{1-\epsilon}\frac{\|\widehat{f}\|_\infty}{\Lambda(\epsilon)}. \tag{3.34}$$

Furthermore, using self-adjointness of G, we have

$$M_{\widehat{f}}^\dagger = (I+\epsilon G)^{-1} M_{\widehat{f}} (I+\epsilon G) = M_{\widehat{f}} + \epsilon (I+\epsilon G)^{-1}(M_{\widehat{f}} G - G M_{\widehat{f}}), \tag{3.35}$$

and thus, since $G1 = 0$,

$$\frac{1}{2}(M_{\widehat{f}} + M_{\widehat{f}}^\dagger)1 = \widehat{f} - \frac{\epsilon}{2}(I+\epsilon G)^{-1} G\widehat{f}.$$

Therefore,

$$\begin{aligned}
\left\|\frac{1}{2}(M_{\widehat{f}} + M_{\widehat{f}}^\dagger)1\right\|_\epsilon^2 &= \left\langle (I - \frac{\epsilon}{2}(I+\epsilon G)^{-1}G)\widehat{f},\ (I+\epsilon G)(I - \frac{\epsilon}{2}(I+\epsilon G)^{-1}G)\widehat{f}\right\rangle \\
&= \left\langle (I - \frac{\epsilon}{2}(I+\epsilon G)^{-1}G)\widehat{f},\ (I + \frac{\epsilon}{2}G)\widehat{f}\right\rangle \\
&\le \left(1 + \frac{\epsilon}{2}\frac{1}{1-\epsilon}\right)\left(1+\frac{\epsilon}{2}\right)\|\widehat{f}\|^2 = \frac{1-\frac{\epsilon^2}{4}}{1-\epsilon}\mathrm{Var}_{\mu_*}[f],
\end{aligned} \tag{3.36}$$

and so

$$v_\pm \le \frac{(1+\epsilon)(1-\frac{\epsilon^2}{4})}{1-\epsilon}\frac{2\,\mathrm{Var}_{\mu_*}[f]}{\Lambda(\epsilon)}. \tag{3.37}$$

Appendix A. Additional proofs. For the sake of completeness and the convenience of the reader, in Appendix A we provide proofs of two results used above that have previously appeared elsewhere in the literature.

First we derive several functional analytic estimates that form an important part of the hypocoercivity method of [15, 16].

Proof of Proposition 2.1. The first property follows from $A1 = A^*1 = 0$ and $\Pi 1 = 1$.

For (2), it easy to verify that $S\Pi = 0$ and taking adjoint gives $\Pi S = 0$.

For (3), note that $T\Pi f = v\nabla_q \Pi f$ and thus $\Pi T\Pi f = \Pi(v\nabla_q \Pi f) = (\nabla_q \Pi f)\Pi v = 0$ (since the velocity v has mean zero).

For (4), note that by (3) we have $\Pi T\Pi = 0$ and thus $B\Pi = 0$. On the other hand, by definition of B we have the identity

$$Bf + (T\Pi)^*(T\Pi)Bf = \Pi T f, \tag{A.1}$$

and thus $\Pi B = B$.

Taking the scalar product of equation (A.1) with Bf and using $\Pi B = B$ and $T\Pi = (I-\Pi)T\Pi$ we obtain

$$\begin{aligned}
\langle Bf,\ Bf\rangle + \langle TBf, TBf\rangle &= \langle -TBf,\ (I-\Pi)f\rangle \\
&\le \|(I-\Pi)f\|\|TBf\| \\
&\le \frac{1}{4}\|(I-\Pi)f\|^2 + \|TBf\|^2.
\end{aligned} \tag{A.2}$$

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

The last inequality gives $\|Bf\| \le \frac{1}{2}\|(I-\Pi)f\|$ while the first inequality gives $\|TBf\| \le \|(I-\Pi)f\|$. □

We end with a derivation of bounds on perturbations to the generator (previously obtained in [7]) that play a key role in proving our new results in Theorems 3.1 and 3.5.

Proof of Lemma 3.2. Let $x \in D(A)$ with $\|x\| = 1$. Define $a = \langle x_0, x\rangle$ so that $\|P^\perp x\|^2 = 1 - |a|^2$ and $|a| \le 1$ with equality if and only if $P^\perp x = 0$. We can decompose $x = ax_0 + \sqrt{1-|a|^2}v$, where: (a) $P^\perp x = 0$, $|a| = 1$, and $v = 0$ or (b) $P^\perp x \neq 0$, $v = P^\perp x/\sqrt{1-|a|^2}$, and $\|v\| = 1$. In either case, $v \perp x_0$.

Using $\langle Mx_0, x_0\rangle = 0$ and $\langle Ax, x\rangle \le -\alpha^{-1}\|P^\perp x\|$ one obtains

$$\begin{aligned}
&\langle (A+\lambda M)x, x\rangle \\
&\le -\alpha^{-1}(1-|a|^2) + 2\lambda a\sqrt{1-|a|^2}\left\langle v, \frac{1}{2}(M+M^\dagger)x_0\right\rangle + \lambda(1-|a|^2)\langle Mv, v\rangle \\
&\le 2\lambda|a|\sqrt{1-|a|^2}V^{1/2} - (1-|a|^2)\left(\alpha^{-1} - \lambda K\right),
\end{aligned}$$

where $V = \left\|\frac{1}{2}(M+M^\dagger)x_0\right\|^2$ and $K = \max\left\{0, \sup_{\|v\|=1}\langle Mv, v\rangle\right\}$. Restricting to $0 \le \lambda < 1/\alpha K$ and using $|a| \le 1$ we can estimate

$$\sup_{x\in D(A), \|x\|=1}\langle (A+\lambda M)x, x\rangle \le \sup_{r\ge 0}\left(2\lambda V^{1/2}r - \left(\alpha^{-1} - \lambda K\right)r^2\right) = \frac{\lambda^2\alpha V}{1-\lambda\alpha K}. \qquad \square$$

Acknowledgment. The second author thanks Gabriel Stoltz and Stefano Olla for useful discussions and suggestions.

References

[1] C. Andrieu, A. Durmus, N. Nüsken, and J. Roussel, *Hypocoercivity of piecewise deterministic Markov process-Monte Carlo*, Ann. Appl. Probab. **31** (2021), no. 5, 2478–2517, DOI 10.1214/20-aap1653. MR4332703

[2] D. Bakry, F. Barthe, P. Cattiaux, and A. Guillin, *A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case*, Electron. Commun. Probab. **13** (2008), 60–66, DOI 10.1214/ECP.v13-1352. MR2386063

[3] J. Bierkens, *Non-reversible Metropolis-Hastings*, Stat. Comput. **26** (2016), no. 6, 1213–1228, DOI 10.1007/s11222-015-9598-x. MR3538633

[4] J. Bierkens, P. Fearnhead, and G. Roberts, *The zig-zag process and super-efficient sampling for Bayesian analysis of big data*, Ann. Statist. **47** (2019), no. 3, 1288–1320, DOI 10.1214/18-AOS1715. MR3911113

[5] J. Bierkens, G. O. Roberts, and P.-A. Zitt, *Ergodicity of the zigzag process*, Ann. Appl. Probab. **29** (2019), no. 4, 2266–2301, DOI 10.1214/18-AAP1453. MR3983339

[6] J. Birrell, M. A. Katsoulakis, and L. Rey-Bellet, *Quantification of model uncertainty on path-space via goal-oriented relative entropy*, ESAIM Math. Model. Numer. Anal. **55** (2021), no. 1, 131–169, DOI 10.1051/m2an/2020070. MR4216833

[7] J. Birrell and L. Rey-Bellet, *Uncertainty quantification for Markov processes via variational principles and functional inequalities*, SIAM/ASA J. Uncertain. Quantif. **8** (2020), no. 2, 539–572, DOI 10.1137/19M1237429. MR4088384

[8] N. Bou-Rabee and J. M. Sanz-Serna, *Randomized Hamiltonian Monte Carlo*, Ann. Appl. Probab. **27** (2017), no. 4, 2159–2194, DOI 10.1214/16-AAP1255. MR3693523

[9] A. Bouchard-Côté, S. J. Vollmer, and A. Doucet, *The bouncy particle sampler: a nonreversible rejection-free Markov chain Monte Carlo method*, J. Amer. Statist. Assoc. **113** (2018), no. 522, 855–867, DOI 10.1080/01621459.2017.1294075. MR3832232

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

[10] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities: a nonasymptotic theory of independence*, with a foreword by Michel Ledoux, Oxford University Press, Oxford, 2013, DOI 10.1093/acprof:oso/9780199535255.001.0001. MR3185193

[11] P. Cattiaux and A. Guillin, *Deviation bounds for additive functionals of Markov processes*, ESAIM Probab. Stat. **12** (2008), 12–29, DOI 10.1051/ps:2007032. MR2367991

[12] K. Chowdhary and P. Dupuis, *Distinguishing and integrating aleatoric and epistemic variation in uncertainty quantification*, ESAIM Math. Model. Numer. Anal. **47** (2013), no. 3, 635–662, DOI 10.1051/m2an/2012038. MR3056403

[13] L. Desvillettes and C. Villani, *On the trend to global equilibrium in spatially inhomogeneous entropy-dissipating systems: the linear Fokker-Planck equation*, Comm. Pure Appl. Math. **54** (2001), no. 1, 1–42, DOI 10.1002/1097-0312(200101)54:1¡1::AID-CPA1¿3.0.CO;2-Q. MR1787105

[14] P. Diaconis, S. Holmes, and R. M. Neal, *Analysis of a nonreversible Markov chain sampler*, Ann. Appl. Probab. **10** (2000), no. 3, 726–752, DOI 10.1214/aoap/1019487508. MR1789978

[15] J. Dolbeault, C. Mouhot, and C. Schmeiser, *Hypocoercivity for kinetic equations with linear relaxation terms* (English, with English and French summaries), C. R. Math. Acad. Sci. Paris **347** (2009), no. 9-10, 511–516, DOI 10.1016/j.crma.2009.02.025. MR2576899

[16] J. Dolbeault, C. Mouhot, and C. Schmeiser, *Hypocoercivity for linear kinetic equations conserving mass*, Trans. Amer. Math. Soc. **367** (2015), no. 6, 3807–3828, DOI 10.1090/S0002-9947-2015-06012-7. MR3324910

[17] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, *Hybrid Monte Carlo*, Phys. Lett. B **195** (1987), no. 2, 216–222, DOI 10.1016/0370-2693(87)91197-x. MR3960671

[18] A. B. Duncan, T. Lelièvre, and G. A. Pavliotis, *Variance reduction using nonreversible Langevin samplers*, J. Stat. Phys. **163** (2016), no. 3, 457–491, DOI 10.1007/s10955-016-1491-2. MR3483241

[19] P. Dupuis and R. S. Ellis, *A weak convergence approach to the theory of large deviations*, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Inc., New York, 1997. A Wiley-Interscience Publication, DOI 10.1002/9781118165904. MR1431744

[20] 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), no. 1, 80–111, DOI 10.1137/15M1025645. MR3455143

[21] P. Dupuis, Y. Liu, N. Plattner, and J. D. Doll, *On the infinite swapping limit for parallel tempering*, Multiscale Model. Simul. **10** (2012), no. 3, 986–1022, DOI 10.1137/110853145. MR3022029

[22] W. E and D. Li, *The Andersen thermostat in molecular dynamics*, Comm. Pure Appl. Math. **61** (2008), no. 1, 96–136, DOI 10.1002/cpa.20198. MR2361305

[23] J.-P. Eckmann and M. Hairer, *Non-equilibrium statistical mechanics of strongly anharmonic chains of oscillators*, Comm. Math. Phys. **212** (2000), no. 1, 105–164, DOI 10.1007/s002200000216. MR1764365

[24] J.-P. Eckmann and M. Hairer, *Spectral properties of hypoelliptic operators*, Comm. Math. Phys. **235** (2003), no. 2, 233–253, DOI 10.1007/s00220-003-0805-9. MR1969727

[25] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet, *Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures*, Comm. Math. Phys. **201** (1999), no. 3, 657–697, DOI 10.1007/s002200050572. MR1685893

[26] J. Fontbona, H. Guérin, and F. Malrieu, *Quantitative estimates for the long-time behavior of an ergodic variant of the telegraph process*, Adv. in Appl. Probab. **44** (2012), no. 4, 977–994. MR3052846

[27] J. Fontbona, H. Guérin, and F. Malrieu, *Long time behavior of telegraph processes under convex potentials*, Stochastic Process. Appl. **126** (2016), no. 10, 3077–3101. MR3542627

[28] F. Gao, A. Guillin, and L. Wu, *Bernstein-type concentration inequalities for symmetric Markov processes*, Theory Probab. Appl. **58** (2014), no. 3, 358–382, DOI 10.1137/S0040585X97986667. MR3403002

[29] K. Gourgoulias, M. A. Katsoulakis, L. Rey-Bellet, and J. Wang, *How biased is your model? Concentration inequalities, information and model bias*, IEEE Trans. Inform. Theory **66** (2020), no. 5, 3079–3097, DOI 10.1109/tit.2020.2977067. MR4089772

[30] A. Guillin, C. Léonard, L. Wu, and N. Yao, *Transportation-information inequalities for Markov processes*, Probab. Theory Related Fields **144** (2009), no. 3-4, 669–695, DOI 10.1007/s00440-008-0159-5. MR2496446

[31] F. Hérau, *Hypocoercivity and exponential time decay for the linear inhomogeneous relaxation Boltzmann equation*, Asymptot. Anal. **46** (2006), no. 3-4, 349–359, DOI 10.3233/asy-2006-741. MR2215889

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

[32] F. Hérau and F. Nier, *Isotropic hypoellipticity and trend to equilibrium for the Fokker-Planck equation with a high-degree potential*, Arch. Ration. Mech. Anal. **171** (2004), no. 2, 151–218, DOI 10.1007/s00205-003-0276-3. MR2034753

[33] C.-R. Hwang, S.-Y. Hwang-Ma, and S.-J. Sheu, *Accelerating diffusions*, Ann. Appl. Probab. **15** (2005), no. 2, 1433–1444, DOI 10.1214/105051605000000025. MR2134109

[34] A. Iacobucci, S. Olla, and G. Stoltz, *Convergence rates for nonequilibrium Langevin dynamics* (English, with English and French summaries), Ann. Math. Qué. **43** (2019), no. 1, 73–98, DOI 10.1007/s40316-017-0091-0. MR3925138

[35] M. A. Katsoulakis, L. Rey-Bellet, and J. Wang, *Scalable information inequalities for uncertainty quantification*, J. Comput. Phys. **336** (2017), 513–545, DOI 10.1016/j.jcp.2017.02.020. MR3622628

[36] P. Lezaud, *Chernoff and Berry-Esséen inequalities for Markov processes*, ESAIM Probab. Statist. **5** (2001), 183–201, DOI 10.1051/ps:2001108. MR1875670

[37] J. C. Mattingly, A. M. Stuart, and D. J. Higham, *Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise*, Stochastic Process. Appl. **101** (2002), no. 2, 185–232, DOI 10.1016/S0304-4149(02)00150-3. MR1931266

[38] P. Monmarché, *Hypocoercive relaxation to equilibrium for some kinetic models*, Kinet. Relat. Models **7** (2014), no. 2, 341–360, DOI 10.3934/krm.2014.7.341. MR3195078

[39] R. M. Neal, *MCMC using Hamiltonian dynamics*, Handbook of Markov chain Monte Carlo, Chapman & Hall/CRC Handb. Mod. Stat. Methods, CRC Press, Boca Raton, FL, 2011, pp. 113–162. MR2858447

[40] F. Peters and G. de With, *Rejection-free Monte Carlo sampling for general potentials*, Phys. Rev. E **85** (2012), 026703.

[41] S. Redon, G. Stoltz, and Z. Trstanova, *Error analysis of modified Langevin dynamics*, J. Stat. Phys. **164** (2016), no. 4, 735–771, DOI 10.1007/s10955-016-1544-6. MR3529154

[42] L. Rey-Bellet and K. Spiliopoulos, *Irreversible Langevin samplers and variance reduction: a large deviations approach*, Nonlinearity **28** (2015), no. 7, 2081–2103, DOI 10.1088/0951-7715/28/7/2081. MR3366637

[43] L. Rey-Bellet and K. Spiliopoulos, *Improving the convergence of reversible samplers*, J. Stat. Phys. **164** (2016), no. 3, 472–494, DOI 10.1007/s10955-016-1565-1. MR3519206

[44] L. Rey-Bellet and L. E. Thomas, *Exponential convergence to non-equilibrium stationary states in classical statistical mechanics*, Comm. Math. Phys. **225** (2002), no. 2, 305–329, DOI 10.1007/s002200100583. MR1889227

[45] G. Stoltz and Z. Trstanova, *Langevin dynamics with general kinetic energies*, Multiscale Model. Simul. **16** (2018), no. 2, 777–806, DOI 10.1137/16M110575X. MR3799045

[46] G. Stoltz and E. Vanden-Eijnden, *Longtime convergence of the temperature-accelerated molecular dynamics method*, Nonlinearity **31** (2018), no. 8, 3748–3769, DOI 10.1088/1361-6544/aac541. MR3826113

[47] C. Villani, *Hypocoercivity*, Mem. Amer. Math. Soc. **202** (2009), no. 950, iv+141, DOI 10.1090/S0065-9266-09-00567-5. MR2562709

[48] C. Wu and C. P. Robert, *Generalized bouncy particle sampler*, 2017, Preprint `arXiv:1706.04781`.

[49] L. Wu, *A deviation inequality for non-reversible Markov processes* (English, with English and French summaries), Ann. Inst. H. Poincaré Probab. Statist. **36** (2000), no. 4, 435–445, DOI 10.1016/S0246-0203(00)00135-7. MR1785390

[50] K. Yosida, *Functional analysis*, reprint of the sixth (1980) edition, Classics in Mathematics, Springer-Verlag, Berlin, 1995, DOI 10.1007/978-3-642-61859-8. MR1336382

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.