Concentration Inequalities and Performance Guarantees for Hypocoercive MCMC Samplers

Jeremiah Birrell * **and Luc Rey-Bellet** *,†

Department of Mathematics and Statistics,
University of Massachusetts Amherst
710 N Pleasant St Amherst, MA, 01002, USA
birrell@math.umass.edu, luc@math.umass.edu

Abstract: In this paper we provide performance guarantees for hypocoercive non-reversible MCMC samplers X_t with invariant measure μ^* and 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 performance guarantees: (a) explicit non-asymptotic confidence intervals for $\int f d\mu^*$ when using a finite time ergodic average with given initial condition μ and (b) uncertainty quantification 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 recent results from Birrell and Rey-Bellet (2018) for coercive dynamics.) The concentration inequality is proved by combining the approach via Feynmann-Kac semigroups first noted by Wu (2000) with the hypocoercive estimates of Dolbeault, Mouhot and Schmeiser (2009, 2015) developed for the Langevin equation and recently generalized to partially deterministic Markov processes by Andrieu et al. (2018).

MSC 2010 subject classifications: 60J25, 47D07, 39B72.
Keywords and phrases: concentration inequality, uncertainty quantification, hypocoercivity, Langevin equation, Bouncy particle sampler, Zig-zag sampler, Hybrid Hamiltonian Monte-Carlo.

1. Introduction and statement of the results

Consider the problem of computing the expected value

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

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 a ergodic stochastic process Q_t with stationary distribution ν^* (in this paper only

*Research supported in part by the National Science Foundation (DMS-1515712) and the Air Force Office of Scientific Research (AFOSR) (FA-9550-18-1-0214).

†Luc Rey-Bellet thanks Gabriel Stoltz and Stefano Olla for useful discussions and suggestions.

continuous-time processes are considered) and use the fact that, by the strong law of large numbers and for suitable f, one has

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f(Q_t)\, dt = \nu^*[f]\,. \tag{2}$$

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 non-reversible 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 Diaconis, Holmes and Neal (2000); Hwang, Hwang-Ma and Sheu (2005); Bierkens (2016); Duncan, Lelièvre and Pavliotis (2016); Rey-Bellet and Spiliopoulos (2016) and many more references therein). In this paper we consider a variety of non-reversible MCMC samplers such as the Langevin equation, and various modifications thereof, as well as partially deterministic Markov processes such as the zig-zag sampler (Bierkens, Fearnhead and Roberts (2019)), the bouncy particle sampler (Peters and de With (2012)), and the hybrid Hamiltonian Monte-Carlo (Duane et al. (1987)). 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 non-reversible MCMC in the extended phase space with an invariant measure $\mu^* = \nu^*\times\rho^*$. The extra dimension $\mathcal{P}$ can be thought 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.

All 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 Desvillettes and Villani (2001); Hérau and Nier (2004); Eckmann and Hairer (2003); Hérau (2006); Villani (2009) it was proved that the Langevin equation is hypocoercive (see also Eckmann, Pillet and Rey-Bellet (1999); Eckmann and Hairer (2000); Rey-Bellet and Thomas (2002); Mattingly, Stuart and Higham (2002) for some earlier and related convergence results). A few years ago Dolbeault, Mouhot and Schmeiser (2009, 2015) found a new, short and very elegant, proof of hypocoercivity, and their techniques have been used for various modifications of the Langevin equation Iacobucci, Olla and Stoltz (2019); Stoltz and Trstanova (2018); Stoltz and Vanden-Eijnden (2018) (some of them without hypoellitpicity) and recently in Andrieu et al. (2018) for a class of partially deterministic Markov processes, among them:

1. The bouncy particle sampler, which was introduced in Peters and de With

(2012) and whose ergodic properties were studied in Bouchard-Côté, Vollmer and Doucet (2018) and Wu and Robert (2017).
2. The zig-zag sampler, introduced in Bierkens, Fearnhead and Roberts (2019) and further studied in Bierkens, Roberts and Zitt (2017), which generalize to higher dimension the so-called telgraph process studied earlier in Fontbona, Guérin and Malrieu (2012, 2016) and Monmarché (2014).
3. The hybrid Hamiltonian Monte Carlo introduced by Duane et al. (1987) and whose ergodic properties are studied in Bou-Rabee and Sanz-Serna (2017), see also E and Li (2008) and Neal (2011).

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 $S^\dagger = S$, $T^\dagger = -T$, 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

$$B = (I + (T\Pi)^*(T\Pi))^{-1}(-T\Pi)^* \tag{3}$$

and the family of modified scalar products $\langle f, g\rangle_\epsilon = \langle f, (I + \epsilon G)g\rangle$, where $G = B + B^\dagger$; 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) this scalar product is equivalent to the $L^2(\mu^*)$-scalar product and it is shown in Dolbeault, Mouhot and Schmeiser (2009, 2015); Andrieu et al. (2018) that, for sufficiently small values of $\epsilon > 0$, the dynamics satisfy a Poincaré inequality for the modified scalar product

$$\langle -Af, f\rangle_\epsilon \geq \Lambda(\epsilon)\,\mathrm{Var}_{\mu^*}(f)\,, \tag{4}$$

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) $f : \mathcal{X} \to \mathbb{R}$:

$$F_T = \frac{1}{T}\int_0^T f(X_t)dt\,. \tag{5}$$

For reversible processes, or more generally for processes whose reversible part satisfies a Poncaré inequality, that is for *coercive* processes, concentration inequalities were obtained first in Lezaud (2001) and then both simplified and greatly generalized in Wu (2000); Cattiaux and Guillin (2008); Guillin et al. (2009); Gao, Guillin and Wu (2014); our approach relies heavily on the ideas developed in these 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

$$\sigma^2(f) \equiv \lim_{T\to\infty} T\,\mathrm{Var}\left(\frac{1}{T}\int_0^T f(X_t)\,dt\right); \tag{6}$$

we take here the alternative point of view of using concentration inequalities, thereby obtaining *non-asymptotic* performance guarantees. In a related manner, it has been advocated in Dupuis et al. (2012) that, to evaluate the performance of a MCMC algorithm, one consider the large deviation rate for the empirical measure itself, an approach used in Rey-Bellet and Spiliopoulos (2015) to analyze non-reversible perturbations of the overdamped Langevin equation.

We prove the following non-asymptotic performance guarantee in Section 2 (see Corollary 1). Here, and in the following, (X_t, P^μ) will denote a $\mathcal{X}$-valued Markov process with initial distribution $X_0 \sim \mu$ (i.e. $(X_0)_* P^\mu = \mu$), E^μ will be the expectation with respect to P^μ, and $\|\cdot\|$ will denote the $L^2(\mu^*)$-norm.

Theorem 1. (Non-asymptotic confidence intervals) *Suppose that the Markov process (X_t, P^μ) satisfies the hypocoercive estimate (4). Then for any bounded observable f, any time $T > 0$, and tolerance level $0 < 1 - \delta < 1$ we have*

$$P^\mu\left(\left|\frac{1}{T}\int_0^T f(X_t)dt - \mu^*[f]\right| \leq r\right) \geq 1 - \delta\,, \tag{7}$$

where

$$r = \sqrt{2v\frac{1}{T}\log\left(\frac{2N}{\delta}\right)} + b\frac{1}{T}\log\left(\frac{2N}{\delta}\right)\,, \tag{8}$$

with

$$v = \frac{(1+\epsilon)(1-\frac{\epsilon^2}{4})}{1-\epsilon}\frac{2\operatorname{Var}_{\mu^*}[f]}{\Lambda(\epsilon)}\,,\ b = \frac{(1+\epsilon)^2}{1-\epsilon}\frac{\|\widehat{f}\|_\infty}{\Lambda(\epsilon)}\,,\ N = \frac{\left\|\frac{d\mu}{d\mu^*}\right\|}{\sqrt{1-\epsilon}}\,. \tag{9}$$

$\Lambda(\epsilon)$ is defined in Eq. (33) and ϵ must satisfy Eq. (34).

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 Birrell and Rey-Bellet (2018), and is based on the general approach to uncertainty quantification developed in Chowdhary and Dupuis (2013); Dupuis et al. (2016); Katsoulakis, Rey-Bellet and Wang (2017); Gourgoulias et al. (2017). In this 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 4).

Theorem 2. (Uncertainty quantification bounds) *Suppose that the baseline Markov process X_t satisfies the hypocoercive estimate (4) and $(\widetilde{X}_t, \widetilde{P}^{\widetilde{\mu}})$ is a stochastic process such that the path space relative entropy satisfies*

$$R\big(\widetilde{P}_T^{\widetilde{\mu}} || P_T^{\mu^*}\big) < \infty\,. \tag{10}$$

Then we have

$$\left| \widetilde{E}^{\widetilde{\mu}} \left[\frac{1}{T} \int_0^T f(\widetilde{X}_t)\, dt \right] - \mu^*[f] \right| \leq \sqrt{2 v \eta_T} + b\eta_T\,, \tag{11}$$

where v and b are given in (9), and

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

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 Dolbeault, Mouhot and Schmeiser (2009, 2015). 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 (4) has been proven by following the strategy of Dolbeault, Mouhot and Schmeiser (2009, 2015). In particular we consider several examples of partially deterministic MCMC samplers studied in the recent paper Andrieu et al. (2018). 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 for $g \in L^2(\nu^*)$

$$\|\nabla_q g\|^2_{L^2(\nu^*)} \geq C_{\nu^*} \operatorname{Var}_{\nu^*}[g]\,. \tag{12}$$

See e.g. Bakry et al. (2008) for conditions on V which imply a Poincaré inequality.

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

$$A = S + T, \text{ with } S = \frac{A + A^\dagger}{2} \text{ and } T = \frac{A - A^\dagger}{2}\,. \tag{13}$$

The following four examples fit within this framework and that will be used to illustrate the utility of our results; see Andrieu et al. (2018) for a proof of hypocoercivity of a more general class of models which covers all examples considered here, as well as Stoltz and Trstanova (2018); Iacobucci, Olla and Stoltz (2019); Stoltz and Vanden-Eijnden (2018) for further examples (some of them being non-equilibrium 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

$$dQ_t = \frac{P_t}{m}dt, \quad dP_t = \left(-\nabla V(Q_t) - \gamma \frac{P_t}{m}\right)dt + \sqrt{\frac{2\gamma}{\beta}}dW_t, \tag{14}$$

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

$$A = \underbrace{\frac{\gamma}{\beta}\Delta_p - \gamma\left(\frac{p}{m}\right)^T \nabla_p}_{=S} + \underbrace{\left(\frac{p}{m}\right)^T \nabla_q - \nabla V(q)^T \nabla_p}_{=T}\,. \tag{15}$$

This is the model originally considered in Dolbeault, Mouhot and Schmeiser (2009, 2015) and several modifications of this models have also been shown to be hypocoercive. For example Redon, Stoltz and Trstanova (2016); Stoltz and Trstanova (2018) consider a Langevin equation with a modified kinetic energy (non-quadratic) so that that ρ^* is not Gaussian and the diffusion needs not be hypoelliptic. Further generalizations of the Langevin equations with general ρ^* are also considered in Andrieu et al. (2018).

2. **(Hybrid Hamiltonian Monte Carlo)** In this randomized version of Hamiltonian Monte-Carlo introduced by Duane et al. (1987), 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

$$A = \underbrace{\lambda(\Pi - I)}_{=S} + \underbrace{\left(\frac{p}{m}\right)^T \nabla_q - \nabla V^T \nabla_p}_{=T}\,. \tag{16}$$

3. **(Bouncy Particle Sampler)** In this sampler, introduced originally in Peters and de With (2012), 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 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
$$R(q)p = p - \frac{p^T \nabla V(q)}{\|\nabla V\|^2}\nabla V\,. \tag{17}$$
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
$$A = \left(\frac{p}{m}\right)^T \nabla_q + \left[\left(\frac{p}{m}\right)^T \nabla V(q)\right]^+ (R - I) + \lambda(\Pi - I)\,, \tag{18}$$
and elementary computations shows that $\mu^* = \nu^* \times \rho^*$ is invariant and
$$S = \left|\left(\frac{p}{m}\right)^T \nabla V(q)\right| (R - I) + \lambda(\Pi - I)\,, \tag{19}$$
$$T = \left(\frac{p}{m}\right)^T \nabla_q + \left(\frac{p}{m}\right)^T \nabla V(q)(R - I)\,. \tag{20}$$
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) i'th component of the velocity is reversed; see Bierkens, Fearnhead and Roberts (2019) for a more detailed discussion. The generator of the Markov process has the form
$$A = v^T\nabla_q + \sum_{i=1}^d [v_i \partial_{q_i} V(q)]^+ (R_i - I) + \lambda(\Pi - I)\,, \tag{21}$$
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
$$S = \sum_{i=1}^d |v_i \partial_{q_i} V(q)| (R_i - I) + \lambda(\Pi - I)\,, \tag{22}$$
$$T = v\nabla_q + \sum_{i=1}^d v_i \partial_{q_i} V(q)(R_i - I)\,. \tag{23}$$

Note that for all the examples considered, it is easy to verify that one has the identity
$$T\Pi = \frac{p}{m}\nabla_q \Pi \tag{24}$$

(with the convention that $p/m = v$ for the zig-zag sampler). This fact is used to establish the following functional analytic estimates (see Dolbeault, Mouhot and Schmeiser (2009, 2015)) which are the basis for the hypocoercive estimates (for the convenience of the reader the proof is in Appendix A.2).

Proposition 1. *Define*

$$B = (I + (T\Pi)^*(T\Pi))^{-1}(-T\Pi)^* \,. \tag{25}$$

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

1. $B1 = B^\dagger 1 = 0$,
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 $L^2(\mu^*)$,

$$\langle f, g\rangle_\epsilon = \langle f, g\rangle + \epsilon\langle f, (B + B^*)g\rangle \,, \;\; \epsilon \in (0,1) \,. \tag{26}$$

As $\|B\| \leq 1/2$, $\langle\cdot,\cdot\rangle_\epsilon$ is an inner product which is equivalent to $\langle\cdot,\cdot\rangle$. As a consequence of Lemma 1 one obtains for suitable f with $\mu^*[f] = 0$:

$$\begin{aligned}\langle Af, f\rangle_\epsilon =& \langle Sf, f\rangle + \epsilon\left[\langle BSf, f\rangle + \langle BTf, f\rangle + \langle SBf, f\rangle - \langle TBf, f\rangle\right] \\ =& \langle (S - \epsilon TB)(I-\Pi)f, (I-\Pi)f\rangle + \epsilon\langle BT\Pi f, \Pi f\rangle \\ &+ \epsilon\left[\langle BS(I-\Pi)f, \Pi f\rangle + \langle BT(I-\Pi)f, \Pi f\rangle\right] \,,\end{aligned} \tag{27}$$

where we have used that $SB = 0$. The various terms in (27) 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^*):

$$\langle f, -Sf\rangle \geq \lambda_p\|(I-\Pi)f\|^2 \tag{28}$$

   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 (24) together with the Poincaré inequality for the measure ν^*, we have

$$\langle f, (T\Pi^\dagger T\Pi)f\rangle = \Pi\left(\frac{p^2}{m^2}\right)\|\nabla_q \Pi f\| \geq \Pi\left(\frac{p^2}{m^2}\right)C_\nu^*\|\Pi f\|^2 \,, \tag{29}$$

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

$$\langle -BT\Pi f, \Pi f\rangle \geq \left(1 - \left(1 + \Pi\left(\frac{p^2}{m^2}\right)C_\nu^*\right)^{-1}\right)\|\Pi f\|^2 \equiv \lambda_q\|\Pi f\|^2 \,. \tag{30}$$

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

$$\|BT(I-\Pi)f\| + \|BS(I-\Pi)f\| \le R_0 \|(I-\Pi)f\| \,. \tag{31}$$

The bound of the first term is the technical part of the proof; for the Langevin equation this is proved in Dolbeault, Mouhot and Schmeiser (2015), and is generalized in Andrieu et al. (2018) for the other samplers (see Lemma 29 and Lemma 32 in particular and the bound in Section 3.3 as well as the bound in Lemma 11 which is specific to the zig-zag sampler).

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

$$\begin{aligned} \langle -Af, f\rangle_\epsilon &\ge \begin{bmatrix} \|\Pi f\| \\ \|(I-\Pi)f\| \end{bmatrix}^T \begin{bmatrix} \epsilon\lambda_q & -\epsilon R_0/2 \\ -\epsilon R_0/2 & \lambda_p - \epsilon \end{bmatrix} \begin{bmatrix} \|\Pi f\| \\ \|(I-\Pi)f\| \end{bmatrix} \\ &\ge \Lambda(\epsilon)\,\mathrm{Var}_{\mu^*}[f]\,, \end{aligned} \tag{32}$$

where

$$\Lambda(\epsilon) \equiv \frac{(\lambda_q - 1)\epsilon + \lambda_p - \sqrt{((\lambda_q+1)\epsilon - \lambda_p)^2 + \epsilon^2 R_0^2}}{2} \tag{33}$$

is the smallest eigenvalue of the matrix in Eq. (32); $\Lambda(\epsilon)$ is positive if

$$0 < \epsilon \le 4\lambda_q\lambda_p/(4\lambda_q + R_0). \tag{34}$$

In the next section, we show how the Poincaré inequality (32) for the modified inner product (26) can be used to derive non-asymptotic confidence intervals and UQ bounds for hypocoercive systems, having in mind the four examples outlined above.

3. Concentration inequalities and performance guarantees via Feynmann-Kac semigroups

In this section, we prove our main new results for hypocoercive systems:

1. A concentration inequality and corresponding non-asymptotic confidence intervals in Section 3.2.
2. UQ bounds in Section 3.3.

The former are obtained by an adaptation of the technique from Wu (2000) and Gao, Guillin and Wu (2014) to hypocoercive systems, which we first summarize.

3.1. Background

As in Wu (2000); Gao, Guillin and Wu (2014), we will prove Bernstein-type concentration inequalities. The following related elementary facts will be used repeatedly (see e.g. the discussion of sub-gamma random variables in Chapter 2 in Boucheron, Lugosi and Massart (2013)):

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

$$\Psi_{v,b}(\lambda) = \frac{\lambda^2 v}{2(1-\lambda b)} \quad \text{for } 0 \le \lambda < 1/b\,. \tag{35}$$

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

$$\Psi^*_{v,b}(r) = \sup_{0\le\lambda<1/b}\{\lambda r - \Psi_{v,b}(\lambda)\} = \frac{2r^2}{v\left(1+\sqrt{1+\frac{2br}{v}}\right)^2} \quad \text{for } r \ge 0 \tag{36}$$

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

$$(\Psi^*_{v,b})^{-1}(\eta) = \inf_{\lambda>0}\left\{\frac{\Psi_{v,b}(\lambda)+\eta}{\lambda}\right\} = \sqrt{2v\eta} + b\eta \quad \text{for } \eta \ge 0\,. \tag{37}$$

Now we summarize the method of Wu (2000); Gao, Guillin and Wu (2014): 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 $\mathcal{X}$ consider the real Hilbert space $L^2(\mu^*)$ with scalar product $\langle\cdot,\cdot\rangle$. We consider the strongly continuous Markov semigroup $\mathcal{P}^V_t : L^2(\mu^*) \to L^2(\mu^*)$ given by

$$\mathcal{P}_t f(x) = E^x[f(X_t)] \tag{38}$$

and whose generator we denote by $(A, D(A))$.

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

$$\mathcal{P}^V_t[f](x) = E^x\left[f(X_t)e^{\int_0^t V(X_s)ds}\right], \tag{39}$$

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

$$\kappa(V) \equiv \sup\{\langle (A+V)g, g\rangle : g \in D(A), \|g\| = 1\} \tag{40}$$

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

$$\langle (A+V-\kappa(V))g\,,\, g\rangle \le 0 \tag{41}$$

and thus by the Lumer-Philipps theorem (see e.g. Chapter IX in Yosida (1995)) the semigroup generated by $A+V-\kappa(V)$ is a contraction semigroup on $L^2(\mu^*)$. This implies that

$$\|\mathcal{P}^V_t\| \le e^{t\kappa(V)}\,, \;\; t \ge 0 \tag{42}$$

(note that Eq. (42) also trivially holds if $\kappa(V)=\infty$). Therefore by the Chernov bound we have

$$\begin{aligned} P^\mu\left(\frac{1}{T}\int_0^T f(X_t)dt-\mu^*[f]>r\right) &\leq \inf_{\lambda>0} e^{-\lambda T r}E^\mu\left[e^{\lambda\int_0^T \widehat{f}(X_t)dt}\right] \\ &\leq \inf_{\lambda>0} e^{-\lambda T r}\int \mathcal{P}_T^{\lambda\widehat{f}}(1)d\mu \\ &\leq \inf_{\lambda>0} e^{-\lambda T r}\left\|\frac{d\mu}{d\mu^*}\right\|\left\|\mathcal{P}_T^{\lambda\widehat{f}}\right\| \\ &\leq \left\|\frac{d\mu}{d\mu^*}\right\| e^{-T\sup_{\lambda>0}\{\lambda r-\kappa(\lambda\widehat{f})\}}. \end{aligned} \tag{43}$$

This basic insight, first noted by Wu (2000), 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 Wu (2000); Lezaud (2001); Cattiaux and Guillin (2008); Guillin et al. (2009); Gao, Guillin and Wu (2014) for many such examples.

3.2. Concentration inequalities

In the hypocoercive examples considered in this paper, the generator is non-reversible 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 for an equivalent modified scalar product. In the following theorem, we show that one still obtains concentration inequalities in this more general setting.

Theorem 3. (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*

$$\|g\|_\#^2\leq\alpha\langle -Ag,g\rangle_\# \quad \text{for all } g\in D(A) \text{ with } \mu^*[g]=0. \tag{44}$$

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

$$P^\mu\left(\pm\left[\frac{1}{T}\int_0^T f(X_t)dt-\mu^*[f]\right]\geq r\right)\leq c^{-1}\left\|\frac{d\mu}{d\mu^*}\right\| e^{-T\Psi^*_{v_\pm,b_\pm}(r)}, \tag{45}$$

where $\Psi^*_{\nu,b}$ *is given in* (36) *and*

$$v_\pm = 2\alpha \left\| \frac{1}{2}(M_{\pm\widehat{f}} + M^\dagger_{\pm\widehat{f}})1 \right\|^2_\# , \qquad b_\pm = \alpha \max\left\{0, \sup_{\|g\|_\#=1} \langle M_{\pm\widehat{f}}g, g\rangle_\# \right\} . \quad (46)$$

We emphasize that $M^\dagger_{\widehat{f}}$ is the adjoint with respect to the $\langle\cdot,\cdot\rangle_\#$-inner product. Also $\nu_\pm$ 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 below).

Proof. The proof is a modification of the strategy used in Gao, Guillin and Wu (2014). We start as in Eq. (43) but use the Lumer-Phillips theorem for the $\|\cdot\|_\#$ norm instead since, by equivalence of the norms, $\mathcal{P}^{\lambda\widehat{f}}_t$ is also a strongly continuous semigroup on $(L^2(\mu^*), \|\cdot\|_\#)$ with the same generator. Using the Chernov bound, the equivalence of the norm and the fact that, by Assumption 2, $\|1\|_\# = \langle 1,1\rangle_\# = \langle 1,1\rangle = 1$, we obtain

$$\begin{aligned}
P^\mu\left(\frac{1}{T}\int_0^T f(X_t)dt \geq \mu^*[f]+r\right) &\leq \inf_{\lambda>0} e^{-\lambda T r} E^\mu\left[e^{\lambda\int_0^T \widehat{f}(X_t)dt}\right] \\
&= \inf_{\lambda>0} e^{-\lambda T r} \int \mathcal{P}^{\lambda\widehat{f}}_T(1)\frac{d\mu}{d\mu^*}d\mu^* \\
&\leq \inf_{\lambda>0} e^{-\lambda T r} \left\|\frac{d\mu}{d\mu^*}\right\| \left\|\mathcal{P}^{\lambda\widehat{f}}_T(1)\right\| \\
&\leq \inf_{\lambda>0} e^{-\lambda T r} \left\|\frac{d\mu}{d\mu^*}\right\| c^{-1} \left\|\mathcal{P}^{\lambda\widehat{f}}_T\right\|_\# \|1\|_\# \\
&\leq c^{-1}\left\|\frac{d\mu}{d\mu^*}\right\| e^{-T\sup_{\lambda>0}(\lambda r - \kappa_\#(\lambda\widehat{f})}, \qquad (47)
\end{aligned}$$

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

$$\kappa_\#(\lambda\widehat{f}) \equiv \sup\left\{\langle(A+\lambda\widehat{f})g,g\rangle_\# : g\in D(A), \|g\|_\# = 1\right\}. \qquad (48)$$

Next we use the following lemma proved in Birrell and Rey-Bellet (2018), which is a generalization of a result in Gao, Guillin and Wu (2014), which itself was a simplification of the argument originally used in Lezaud (2001). For completeness, the proof is given in the appendix.

Lemma 1. *Let H be a real Hilbert space, $A : D(A) \subset H \to H$ a linear operator, and $M : H \to H$ a bounded linear operator. Assume there exists $\alpha > 0$ and $x_0 \in H$ with $\|x_0\| = 1$ such that*

$$\langle Mx_0, x_0\rangle = 0 \quad \textit{and} \quad \langle Ax, x\rangle \leq -\alpha^{-1}\|P^\perp x\|^2 \qquad (49)$$

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

$$\sup_{x\in D(A),\|x\|=1} \langle (A+\lambda M)x, x\rangle \leq \frac{\lambda^2 \alpha V}{1-\lambda\alpha K} = \Psi_{2\alpha V,\alpha K}(\lambda) \tag{50}$$

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

$$V = \left\| \frac{1}{2}(M+M^\dagger)x_0 \right\|^2, \quad K = \max\left\{0, \sup_{\|y\|=1} \langle My, y\rangle \right\}. \tag{51}$$

To use this result we take $x_0 = 1$, A to be the generator on $(L^2(\mu^*), \|\cdot\|_\#)$, and $M = M_{\widehat{f}}$. 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

$$\langle Ag, 1\rangle_\# = \langle Ag, 1\rangle = 0, \;\; g \in D(A). \tag{52}$$

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 1 to obtain

$$\kappa_\#(\lambda\widehat{f}) = \sup_{g\in D(A),\|g\|_\#=1} \langle (A+\lambda\widehat{f})g, g\rangle_\# \leq \Psi_{v_+,b_+}(\lambda) \tag{53}$$

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

$$v_+ = 2\alpha \left\| \frac{1}{2}(M_{\widehat{f}} + M_{\widehat{f}}^\dagger)1 \right\|_\#^2, \;\; b_+ = \alpha \max\left\{0, \sup_{\|g\|_\#=1} \langle M_{\widehat{f}}g, g\rangle_\# \right\} \tag{54}$$

(as was given in (46)). Therefore

$$\begin{aligned} P^\mu\left(\frac{1}{T}\int_0^T f(X_t)dt \geq \mu^*[f] + r\right) &\leq c^{-1}\left\|\frac{d\mu}{d\mu^*}\right\| e^{-T\sup_{0\leq\lambda<1/b_+}\{\lambda r - \Psi_{v_+,b_+}(\lambda)\}} \\ &= c^{-1}\left\|\frac{d\mu}{d\mu^*}\right\| e^{-T\Psi^*_{v_+,b_+}(r)}. \end{aligned} \tag{55}$$

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

As an immediate corollary we obtain a non-asymptotic confidence interval.

Corollary 1. (Confidence intervals). *Under the same assumptions as in Theorem 3, given a time T and a confidence level $0 < 1-\delta < 1$ we have*

$$P^\mu\left(\frac{1}{T}\int_0^T f(X_t)dt - \mu^*[f] \in (-r_-, r_+)\right) \geq 1-\delta \tag{56}$$

where

$$r_{\pm} = \sqrt{2v_{\pm}\frac{1}{T}\log\left(\frac{2N}{\delta}\right)} + b_{\pm}\frac{1}{T}\log\left(\frac{2N}{\delta}\right), \tag{57}$$

with $N = c^{-1}\left\|\frac{d\mu}{d\mu^*}\right\|$ *and* $v_{\pm}$ *and* $b_{\pm}$ *given in* (46).

Proof. Define $\eta = \frac{1}{T}\log\left(\frac{2N}{\delta}\right)$ (note that $N \geq 1$ follows from Assumptions 1 and 2 of Theorem 3), so that $r_{\pm} = (\Psi^*_{v_{\pm},b_{\pm}})^{-1}(\eta)$, with $r_{\pm}$ given as in (57). Using $r = r_{\pm}$ in the concentration bound in Theorem 3 we find

$$P^{\mu}\left(\pm\left[\frac{1}{T}\int_0^T f(X_t)dt - \mu^*[f]\right] \geq r_{\pm}\right) \leq Ne^{-T\eta} = \frac{\delta}{2}. \tag{58}$$

The result (56) then follows from a union bound. □

3.3. *Robustness bounds on steady state bias due to model form uncertainty*

Following on the recent results in Gourgoulias et al. (2017); Birrell and Rey-Bellet (2018), we also use concentration inequalities 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 1. *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 processes.*

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:

$$R\big(\widetilde{P}^{\widetilde{\mu}}_T \| P^{\mu}_T\big) < \infty\,. \tag{59}$$

See the supplementary material to Dupuis et al. (2016) for a collection of techniques that can be used to bound the path-space relative entropy (59) for various classes of alternative models.

Given an observable f we consider the ergodic averages

$$\widetilde{F}_T = \frac{1}{T}\int_0^T f(\widetilde{X}_t)dt\,, \quad F_T = \frac{1}{T}\int_0^T f(X_t)dt\,, \tag{60}$$

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

$$\textbf{Bias:} \quad \widetilde{E}^{\widetilde{\mu}}[\widetilde{F}_T] - E^{\mu}[F_T]\,. \tag{61}$$

Theorem 4. (Uncertainty Quantification bounds). *Let (X_t, P^x), $x \in \mathcal{X}$, be a family of Markov process satisfying the assumptions of Theorem 3, μ be an initial distribution, and $(X_t, \widetilde{P}^{\widetilde{\mu}})$ 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 (46) 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

$$\eta_\infty = \lim_{T\to\infty}\frac{1}{T}R\big(\widetilde{P}_T^{\widetilde{\mu}}||P_T^{\mu^*}\big) \tag{62}$$

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

$$\pm\left(\widetilde{\mu}^*[f] - \mu^*[f]\right) \leq \sqrt{2v_{\pm}\eta_\infty} + b_{\pm}\eta_\infty\,. \tag{63}$$

Proof. The proof proceeds along the same line as in Birrell and Rey-Bellet (2018) to which we refer for more details. The starting point is the Gibbs information inequality Chowdhary and Dupuis (2013); Dupuis et al. (2016): for g bounded and measurable and probability measures Q and $\widetilde{Q}$

$$\pm\left(E_{\widetilde{Q}}E[g] - E_Q[g]\right) \leq \inf_{\lambda>0}\left\{\frac{\log E_Q[e^{\pm\lambda(g-E_Q[g])}] + R(\widetilde{Q}||Q)}{\lambda}\right\}\,. \tag{64}$$

This is a direct consequence of the Gibbs variational principle for the relative entropy, Dupuis and Ellis (1997).

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) and then divide both sides by T:

$$\begin{aligned}
&\pm\left(\widetilde{E}^{\widetilde{\mu}}[\widetilde{F}_T] - E^{\mu}[F_T]\right)\\
&\leq \inf_{\lambda>0}\left\{\frac{\log E^{\mu}[e^{\pm\lambda T(F_T - E^{\mu}[F_T])}] + R(\widetilde{P}_T^{\widetilde{\mu}}||P_T^{\mu})}{\lambda T}\right\}\\
&\leq \underbrace{\inf_{\lambda>0}\left\{\frac{\log E^{\mu}[e^{\pm\lambda T(F_T-\mu^*[f])}] + R(\widetilde{P}_T^{\widetilde{\mu}}||P_T^{\mu})}{\lambda T}\right\}}_{=(\mathrm{I})} \mp \underbrace{\left(E^{\mu}[F_T] - \mu^*[f]\right)}_{=(\mathrm{II})}\,.
\end{aligned} \tag{65}$$

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

$$\begin{aligned}|(\mathrm{II})| = \left|E^\mu\left[\frac{1}{T}\int_0^T \widehat{f}(X_t)\,dt\right]\right| &\leq \frac{1}{T}\int_0^T \left|\left\langle \frac{d\mu}{d\mu^*}, \mathcal{P}_t\widehat{f}\right\rangle\right| dt \\ &\leq \frac{1}{T}\int_0^T e^{-t/\alpha}\left\|\frac{d\mu}{d\mu^*}\right\| \frac{C}{c}\|\widehat{f}\|dt \\ &= \frac{C}{c}\frac{1-e^{-T/\alpha}}{T/\alpha}\left\|\frac{d\mu}{d\mu^*}\right\|\sqrt{\mathrm{Var}_{\mu^*}[f]}\,.\end{aligned} \tag{66}$$

To bound the term (I), we use Lemma 1 to bound the moment generating function, similarly to the proof of Theorem 3:

$$\begin{aligned}(\mathrm{I}) &= \inf_{\lambda>0}\left\{\frac{\log\int \mathcal{P}_T^{\pm\lambda\widehat{f}}(1)d\mu + R\big(\widetilde{P}_T^{\widetilde{\mu}}||P_T^\mu\big)}{\lambda T}\right\} \\ &\leq \inf_{\lambda>0}\left\{\frac{\log\left(c^{-1}\left\|\frac{d\mu}{d\mu^*}\right\|e^{T\kappa_\#(\pm\lambda\widehat{f})}\right) + R\big(\widetilde{P}_T^{\widetilde{\mu}}||P_T^\mu\big)}{\lambda T}\right\} \\ &= \inf_{\lambda>0}\left\{\frac{\kappa_\#(\pm\lambda\widehat{f})+\eta_T}{\lambda}\right\} \\ &\leq \inf_{\lambda>0}\left\{\frac{\Psi_{v_\pm,b_\pm}(\lambda)+\eta_T}{\lambda}\right\} \\ &= (\Psi^*_{v_\pm,b_\pm})^{-1}(\eta_T) = \sqrt{2v_\pm\eta_T} + b_\pm\eta_T\,.\end{aligned} \tag{67}$$

Finally, by taking $T\to\infty$ we obtain the bounds in Eq. (63) $\square$

3.4. Application to hypocoercive samplers

Theorems 1 and 2 for hypocoercive MCMC samplers follow rather immediately from Corollary 1 and from Theorem 4. We first verify the three assumptions in Theorem 3. The modified scalar product is $\langle f,g\rangle_\epsilon = \langle f,g\rangle + \epsilon\langle f,Gg\rangle$ with $G1=0$ and $\|G\|\leq 1$. Therefore we we have $c=(1-\epsilon)^{1/2}$, $C=(1+\epsilon)^{1/2}$, and $\langle f\,,\,1\rangle_\epsilon = \langle f\,,\,1\rangle$, and, for $\epsilon>0$ sufficiently small (see Eq. (34)), by Eq. (32) we have $\alpha = \frac{1+\epsilon}{\Lambda(\epsilon)}$.

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

$$b_\pm = \alpha\max\left\{0, \sup_{\|g\|_\epsilon=1}\langle M_{\pm\widehat{f}}g,g\rangle_\epsilon\right\} \leq \frac{(1+\epsilon)^2}{1-\epsilon}\frac{\|\widehat{f}\|_\infty}{\Lambda(\epsilon)}\,. \tag{68}$$

Furthermore, using self-adjointness of G, we have

$$M_{\widehat{f}}^\dagger = (I+\epsilon G)^{-1}M_{\widehat{f}}(I+\epsilon G) = M_{\widehat{f}} + \epsilon(I+\epsilon G)^{-1}(M_{\widehat{f}}G - GM_{\widehat{f}}), \tag{69}$$

and thus, since $G1 = 0$,

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

Therefore

$$\begin{aligned}
\left\| \frac{1}{2}(M_{\widehat{f}} + M^*_{\widehat{f}})1 \right\|_\epsilon^2 &= \left\langle (I - \frac{\epsilon}{2}(I+\epsilon G)^{-1}G)\widehat{f} \, , \, (I+\epsilon G)(I - \frac{\epsilon}{2}(I+\epsilon G)^{-1}G)\widehat{f} \right\rangle \\
&= \left\langle (I - \frac{\epsilon}{2}(I+\epsilon G)^{-1}G)\widehat{f} \, , \, (I + \frac{\epsilon}{2}G)\widehat{f} \right\rangle \\
&\leq \left(1 + \frac{\epsilon}{2}\frac{1}{1-\epsilon}\right)\left(1 + \frac{\epsilon}{2}\right) \|\widehat{f}\|^2 = \frac{1 - \frac{\epsilon^2}{4}}{1-\epsilon} \mathrm{Var}_{\mu^*}[f] \, , \qquad (70)
\end{aligned}$$

and so

$$v_\pm \leq \frac{(1+\epsilon)(1-\frac{\epsilon^2}{4})}{1-\epsilon} \frac{2\,\mathrm{Var}_{\mu^*}[f]}{\Lambda(\epsilon)} \, . \qquad (71)$$

Appendix A: Some proofs

A.1. Proof of Lemma 1

Proof. 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 $|a| \leq 1$ with equality if and only if $P^\perp x = 0$. We can decompose $x = a x_0 + \sqrt{1 - |a|^2} v$, 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 $\langle M x_0, x_0 \rangle = 0$ and $\langle Ax, x \rangle \leq -\alpha^{-1} \|P^\perp x\|$ we obtain

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

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 $0 \leq \lambda < 1/\alpha K$ and using $|a| \leq 1$ we can estimate

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

□

A.2. Proof of Proposition 1

Proof. The first property follows from $A1 = A^\dagger 1 = 0$ 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

$$Bf + (T\Pi)^*(T\Pi)Bf = \Pi T f\,, \tag{72}$$

and thus $\Pi B = B$.

Taking the scalar product of Eq. (72) with Bf and using $\Pi B = B$ and $T\Pi = (I - \Pi)T\Pi$ we obtain

$$\begin{aligned} \langle Bf\,,\,Bf\rangle + \langle TBf, TBf\rangle &= \langle -TBf\,,\,(I-\Pi)f\rangle \\ &\le \|(I-\Pi)f\|\|TBf\| \\ &\le \frac{1}{4}\|(I-\Pi)f\|^2 + \|TBf\|^2\,. \end{aligned} \tag{73}$$

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

References

Andrieu, C., Durmus, A., Nüsken, N. and Roussel, J. (2018). Hypercoercivity of Piecewise Deterministic Markov Process-Monte Carlo. *ArXiv e-prints* arXiv:1808.08592.

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

Bierkens, J. (2016). Non-reversible Metropolis-Hastings. *Stat. Comput.* **26** 1213–1228. MR3538633

Bierkens, J., Fearnhead, P. and Roberts, G. (2019). The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.* **47** 1288–1320. MR3911113

Bierkens, J., Roberts, G. and Zitt, P.-A. (2017). Ergodicity of the zigzag process. *arXiv e-prints* arXiv:1712.09875.

Birrell, J. and Rey-Bellet, L. (2018). Uncertainty Quantification for Markov Processes via Variational Principles and Functional Inequalities. *arXiv e-prints* arXiv:1812.05174.

Bou-Rabee, N. and Sanz-Serna, J. M. (2017). Randomized Hamiltonian Monte Carlo. *Ann. Appl. Probab.* **27** 2159–2194. MR3693523

Bouchard-Côté, A., Vollmer, S. J. and Doucet, A. (2018). The bouncy particle sampler: a nonreversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.* **113** 855–867. MR3832232

Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration inequalities*. Oxford University Press, Oxford A nonasymptotic theory of independence, With a foreword by Michel Ledoux. MR3185193

Cattiaux, P. and Guillin, A. (2008). Deviation bounds for additive functionals of Markov processes. *ESAIM Probab. Stat.* **12** 12–29. MR2367991

Chowdhary, K. and Dupuis, P. (2013). Distinguishing and integrating aleatoric and epistemic variation in uncertainty quantification. *ESAIM Math. Model. Numer. Anal.* **47** 635–662. MR3056403

Desvillettes, L. and Villani, C. (2001). On the trend to global equilibrium in spatially inhomogeneous entropy-dissipating systems: the linear Fokker-Planck equation. *Comm. Pure Appl. Math.* **54** 1–42. MR1787105

Diaconis, P., Holmes, S. and Neal, R. M. (2000). Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.* **10** 726–752. MR1789978

Dolbeault, J., Mouhot, C. and Schmeiser, C. (2009). Hypocoercivity for kinetic equations with linear relaxation terms. *C. R. Math. Acad. Sci. Paris* **347** 511–516. MR2576899

Dolbeault, J., Mouhot, C. and Schmeiser, C. (2015). Hypocoercivity for linear kinetic equations conserving mass. *Trans. Amer. Math. Soc.* **367** 3807–3828. MR3324910

Duane, S., Kennedy, A. D., Pendleton, B. J. and Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B* **195** 216 - 222.

Duncan, A. B., Lelièvre, T. and Pavliotis, G. A. (2016). Variance reduction using nonreversible Langevin samplers. *J. Stat. Phys.* **163** 457–491. MR3483241

Dupuis, P. and Ellis, R. S. (1997). *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 A Wiley-Interscience Publication. MR1431744

Dupuis, P., Liu, Y., Plattner, N. and Doll, J. D. (2012). On the infinite swapping limit for parallel tempering. *Multiscale Model. Simul.* **10** 986–1022. MR3022029

Dupuis, P., Katsoulakis, M. A., Pantazis, Y. and Plecháč, P. (2016). Path-space information bounds for uncertainty quantification and sensitivity analysis of stochastic dynamics. *SIAM/ASA J. Uncertain. Quantif.* **4** 80–111. MR3455143

E, W. and Li, D. (2008). The Andersen thermostat in molecular dynamics. *Comm. Pure Appl. Math.* **61** 96–136. MR2361305

Eckmann, J.-P. and Hairer, M. (2000). Non-equilibrium statistical mechanics of strongly anharmonic chains of oscillators. *Comm. Math. Phys.* **212** 105–164. MR1764365

Eckmann, J.-P. and Hairer, M. (2003). Spectral properties of hypoelliptic operators. *Comm. Math. Phys.* **235** 233–253. MR1969727

Eckmann, J.-P., Pillet, C.-A. and Rey-Bellet, L. (1999). Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures. *Comm. Math. Phys.* **201** 657–697. MR1685893

Fontbona, J., Guérin, H. and Malrieu, F. (2012). Quantitative estimates for the long-time behavior of an ergodic variant of the telegraph process. *Adv. in Appl. Probab.* **44** 977–994. MR3052846

Fontbona, J., Guérin, H. and Malrieu, F. (2016). Long time behavior of telegraph processes under convex potentials. *Stochastic Process. Appl.* **126** 3077–3101. MR3542627

Gao, F., Guillin, A. and Wu, L. (2014). Bernstein-type concentration inequalities for symmetric Markov processes. *Theory Probab. Appl.* **58** 358–382. MR3403002

Gourgoulias, K., Katsoulakis, M. A., Rey-Bellet, L. and Wang, J. (2017). How biased is your model? Concentration Inequalities, Information and Model Bias. *CoRR* **abs/1706.10260**.

Guillin, A., Léonard, C., Wu, L. and Yao, N. (2009). Transportation-information inequalities for Markov processes. *Probab. Theory Related Fields* **144** 669–695. MR2496446

Hérau, F. (2006). Hypocoercivity and exponential time decay for the linear inhomogeneous relaxation Boltzmann equation. *Asymptot. Anal.* **46** 349–359. MR2215889

Hérau, F. and Nier, F. (2004). Isotropic hypoellipticity and trend to equilibrium for the Fokker-Planck equation with a high-degree potential. *Arch. Ration. Mech. Anal.* **171** 151–218. MR2034753

Hwang, C.-R., Hwang-Ma, S.-Y. and Sheu, S.-J. (2005). Accelerating diffusions. *Ann. Appl. Probab.* **15** 1433–1444. MR2134109

Iacobucci, A., Olla, S. and Stoltz, G. (2019). Convergence rates for nonequilibrium Langevin dynamics. *Ann. Math. Qué.* **43** 73–98. MR3925138

Katsoulakis, M. A., Rey-Bellet, L. and Wang, J. (2017). Scalable information inequalities for uncertainty quantification. *J. Comput. Phys.* **336** 513–545. MR3622628

Lezaud, P. (2001). Chernoff and Berry-Esséen inequalities for Markov processes. *ESAIM Probab. Statist.* **5** 183–201. MR1875670

Mattingly, J., Stuart, A. and Higham, D. (2002). Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.* **101** 185–232. MR1931266

Monmarché, P. (2014). Hypocoercive relaxation to equilibrium for some kinetic models. *Kinet. Relat. Models* **7** 341–360. MR3195078

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

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

Redon, S., Stoltz, G. and Trstanova, Z. (2016). Error analysis of modified Langevin dynamics. *J. Stat. Phys.* **164** 735–771. MR3529154

Rey-Bellet, L. and Spiliopoulos, K. (2015). Irreversible Langevin samplers and variance reduction: a large deviations approach. *Nonlinearity* **28** 2081–2103. MR3366637

Rey-Bellet, L. and Spiliopoulos, K. (2016). Improving the convergence of reversible samplers. *J. Stat. Phys.* **164** 472–494. MR3519206

Rey-Bellet, L. and Thomas, L. E. (2002). Exponential convergence to non-equilibrium stationary states in classical statistical mechanics. *Comm. Math. Phys.* **225** 305–329. MR1889227

Stoltz, G. and Trstanova, Z. (2018). Langevin dynamics with general kinetic energies. *Multiscale Model. Simul.* **16** 777–806. MR3799045

Stoltz, G. and Vanden-Eijnden, E. (2018). Longtime convergence of the temperature-accelerated molecular dynamics method. *Nonlinearity* **31** 3748–3769. MR3826113

Villani, C. (2009). Hypocoercivity. *Mem. Amer. Math. Soc.* **202** iv+141. MR2562709

Wu, L. (2000). A deviation inequality for non-reversible Markov processes. *Ann. Inst. H. Poincaré Probab. Statist.* **36** 435–445. MR1785390

Wu, C. and Robert, C. P. (2017). Generalized Bouncy Particle Sampler. *arXiv e-prints* arXiv:1706.04781.

Yosida, K. (1995). *Functional analysis. Classics in Mathematics*. Springer-Verlag, Berlin Reprint of the sixth (1980) edition. MR1336382