Submitted to Bernoulli
arXiv: stat.ME/1410.5110

The Geometric Foundations of Hamiltonian Monte Carlo

Michael Betancourt, Simon Byrne, Sam Livingstone, and Mark Girolami

Michael Betancourt is a Postdoctoral Research Associate at the University of Warwick, Coventry CV4 7AL, UK E-mail: betanalpha@gmail.com. *Simon Byrne is an EPSRC Postdoctoral Research Fellow at University College London, Gower Street, London, WC1E 6BT. Sam Livingstone is a PhD candidate at University College London, Gower Street, London, WC1E 6BT. Mark Girolami is an ESPRC Established Career Research Fellow at the University of Warwick, Coventry CV4 7AL, UK.*

Although Hamiltonian Monte Carlo has proven an empirical success, the lack of a rigorous theoretical understanding of the algorithm has in many ways impeded both principled developments of the method and use of the algorithm in practice. In this paper we develop the formal foundations of the algorithm through the construction of measures on smooth manifolds, and demonstrate how the theory naturally identifies efficient implementations and motivates promising generalizations.

Keywords: Markov Chain Monte Carlo, Hamiltonian Monte Carlo, Disintegration, Differential Geometry, Smooth Manifold, Fiber Bundle, Riemannian Geometry, Symplectic Geometry.

The frontier of Bayesian inference requires algorithms capable of fitting complex models with hundreds, if not thousands of parameters, intricately bound together with nonlinear and often hierarchical correlations. Hamiltonian Monte Carlo (Duane et al., 1987; Neal, 2011) has proven tremendously successful at extracting inferences from these models, with applications spanning computer science (Sutherland, Póczos and Schneider, 2013; Tang, Srivastava and Salakhutdinov, 2013), ecology (Schofield et al., 2014; Terada, Inoue and Nishihara, 2013), epidemiology (Cowling et al., 2012), linguistics (Husain, Vasishth and Srinivasan, 2014), pharmacokinetics (Weber et al., 2014), physics (Jasche et al., 2010; Porter and Carré, 2014; Sanders, Betancourt and Soderberg, 2014; Wang et al., 2014), and political science (Ghitza and Gelman, 2014), to name a few. Despite such widespread empirical success, however, there remains an air of mystery concerning the efficacy of the algorithm.

This lack of understanding not only limits the adoption of Hamiltonian Monte Carlo but may also foster fragile implementations that restrict the scalability of the algorithm. Consider, for example, the Compressible Generalized Hybrid Monte Carlo scheme of Fang, Sanz-Serna and Skeel (2014) and the particular implementation in Lagrangian Dynamical Monte Carlo (Lan et al., 2012). In an effort to reduce the computational burden of the algorithm, the authors sacrifice the costly volume-preserving numerical integrators typical of Hamiltonian Monte Carlo. Although this leads to improved performance in some low-dimensional models, the performance rapidly diminishes with increasing model dimension (Lan et al., 2012) in sharp contrast to standard Hamiltonian Monte Carlo. Clearly, the volume-preserving numerical integrator is somehow critical to scalable performance; but why?

In this paper we develop the theoretical foundation of Hamiltonian Monte Carlo in order to answer questions like these. We demonstrate how a formal understanding naturally identifies the properties critical to the success of the algorithm, hence immediately providing a framework for

robust implementations. Moreover, we discuss how the theory motivates several generalizations that may extend the success of Hamiltonian Monte Carlo to an even broader array of applications.

We begin by considering the properties of efficient Markov kernels and possible strategies for constructing those kernels. This construction motivates the use of tools in differential geometry, and we continue by curating a coherent theory of probabilistic measures on smooth manifolds. In the penultimate section we show how that theory provides a skeleton for the development, implementation, and formal analysis of Hamiltonian Monte Carlo. Finally, we discuss how this formal perspective directs generalizations of the algorithm.

Without a familiarity with differential geometry a complete understanding of this work will be a challenge, and we recommend that readers without a background in the subject only scan through Section 2 to develop some intuition for the probabilistic interpretation of forms, fiber bundles, Riemannian metrics, and symplectic forms, as well as the utility of Hamiltonian flows. For those readers interesting in developing new implementations of Hamiltonian Monte Carlo we recommend a more careful reading of these sections and suggest introductory literature on the mathematics necessary to do so in the introduction of Section 2.

1. Constructing Efficient Markov Kernels

Bayesian inference is conceptually straightforward: the information about a system is first modeled with the construction of a posterior distribution, and then statistical questions can be answered by computing expectations with respect to that distribution. Many of the limitations of Bayesian inference arise not in the modeling of a posterior distribution but rather in computing the subsequent expectations. Because it provides a generic means of estimating these expectations, Markov Chain Monte Carlo has been critical to the success of the Bayesian methodology in practice.

In this section we first review the Markov kernels intrinsic to Markov Chain Monte Carlo and then consider the dynamic systems perspective to motivate a strategy for constructing Markov kernels that yield computationally efficient inferences.

1.1. Markov Kernels

Consider a probability space,

$$(Q, \mathcal{B}(Q), \pi),$$

with an n-dimensional sample space, Q, the Borel σ-algebra over Q, $\mathcal{B}(Q)$, and a distinguished probability measure, π. In a Bayesian application, for example, the distinguished measure would be the posterior distribution and our ultimate goal would be the estimation of expectations with respect to the posterior, $\mathbb{E}_{\pi}[f]$.

A *Markov kernel*, τ, is a map from an element of the sample space and the σ-algebra to a probability,

$$\tau : Q \times \mathcal{B}(Q) \to [0, 1],$$

such that the kernel is a measurable function in the first argument,

$$\tau(\cdot, A) : Q \to [0, 1], \, \forall A \in \mathcal{B}(Q),$$

and a probability measure in the second argument,

$$\tau(q, \cdot) : \mathcal{B}(Q) \to [0, 1], \, \forall q \in Q.$$

By construction the kernel defines a map,

$$\tau : Q \to \mathcal{P}(Q) \,,$$

where $\mathcal{P}(Q)$ is the space of probability measures over Q; intuitively, at each point in the sample space the kernel defines a probability measure describing how to sample a new point.

By averaging the Markov kernel over all initial points in the state space we can construct a *Markov transition* from probability measures to probability measures,

$$\mathcal{T} : \mathcal{P}(Q) \to \mathcal{P}(Q) \,,$$

by

$$\pi'(A) = \pi\mathcal{T}(A) = \int \tau(q, A) \, \pi(\mathrm{d}q) \,, \, \forall q \in Q, \, A \in \mathcal{B}(Q) \,.$$

When the transition is aperiodic, irreducible, Harris recurrent, and preserves the target measure, $\pi\mathcal{T} = \pi$, its repeated application generates a *Markov chain* that will eventually explore the entirety of π. Any realization, $(q_0, q_1, \ldots, q_N)$, of the Markov chain yields *Markov Chain Monte Carlo estimators* of any expectation (Roberts et al., 2004; Meyn and Tweedie, 2009). Formally, for any integrable function $f \in L^1(Q, \pi)$ we can construct estimators,

$$\hat{f}_N(q_0) = \frac{1}{N} \sum_{n=0}^{N} f(q_n) \,,$$

that are consistent for any initial $q_0 \in Q$,

$$\lim_{N \to \infty} \hat{f}_N(q_0) \xrightarrow{\mathcal{P}} \mathbb{E}_\pi[f] \,.$$

Here δ_q is the *Dirac measure* that concentrates on q,

$$\delta_q(A) \propto \begin{cases} 0, & q \notin A \\ 1, & q \in A \end{cases} \,, \, q \in Q, A \in \mathcal{B}(Q) \,.$$

In practice we are interested not just in Markov chains that explore the target distribution as $N \to \infty$ but in Markov chains that can explore and yield precise Markov Chain Monte Carlo estimators in only a finite number of transitions. From this perspective the efficiency of a Markov chain can be quantified in terms of the *autocorrelation*, which measures the dependence of any square integrable test function, $f \in L^2(Q, \pi)$, before and after the application of the Markov transition

$$\rho[f] \equiv \frac{\int f(q_1) \, f(q_2) \, \tau(q_1, \mathrm{d}q_2) \, \pi(\mathrm{d}q_1) - \int f(q_2) \, \pi(\mathrm{d}q_2) \int f(q_1) \, \pi(\mathrm{d}q_1)}{\int f^2(q) \, \pi(\mathrm{d}q) - \left(\int f(q) \, \pi(\mathrm{d}q) \right)^2} .$$

In the best case the Markov kernel reproduces the target measure,

$$\tau(q, A) = \pi(A) \,, \, \forall q \in Q,$$

and the autocorrelation vanishes for all test functions, $\rho[f] = 0$. Alternatively, a Markov kernel restricted to a Dirac measure at the initial point,

$$\tau(q, A) = \delta_q(A) \,,$$

moves nowhere and the autocorrelations are maximal for any test function, $\rho[f] = 1$. Note that we are disregarding anti-autocorrelated chains, whose performance is highly sensitive to the particular f under consideration.

Given a target measure, any Markov kernel will lie in between these two extremes; the more of the target measure a kernel explores the smaller the autocorrelations, while the more localized the exploration to the initial point the larger the autocorrelations. Unfortunately, common Markov kernels like *Gaussian Random walk Metropolis* (Robert and Casella, 1999) and the *Gibbs sampler* (Geman and Geman, 1984; Gelfand and Smith, 1990) degenerate into local exploration, and poor efficiency, when targeting the complex distributions of interest. Even in two-dimensions, for example, nonlinear correlations in the target distribution constrain the n-step transition kernels to small neighborhoods around the initial point (Figure 1).

In order for Markov Chain Monte Carlo to perform well on these contemporary problems we need to be able to engineer Markov kernels that maintain exploration, and hence small autocorrelations, when targeting intricate distributions. The construction of such kernels is greatly eased with the use of measure-preserving maps.

1.2. Markov Kernels Induced From Measure-Preserving Maps

Directly constructing a Markov kernel that targets π, let alone an efficient Markov kernel, can be difficult. Instead of constructing a kernel directly, however, we can construct one indirectly be defining a family of measure-preserving maps (Petersen, 1989).

Formally, let Γ be some space of continuous, bijective maps from the space into itself,

$$t : Q \to Q, \forall t \in \Gamma,$$

that each preserves the target measure,

$$t_* \pi = \pi,$$

where the pushforward measure, $t_* \pi$, is defined as

$$(t_* \pi)(A) \equiv \left(\pi \circ t^{-1}\right)(A), \forall A \in \mathcal{B}(Q).$$

Given a σ-algebra, $\mathcal{G}$, over Γ, the choice of a probability measure over $\mathcal{G}$, γ, defines a probability space,

$$(\Gamma, \mathcal{G}, \gamma),$$

which induces a Markov kernel as an iterated random function (Diaconis and Freedman, 1999),

$$\tau(q, A) \equiv \int_{\Gamma} \gamma(\mathrm{d}t) \, \mathbb{I}_A(t\,(q)), \tag{1}$$

where $\mathbb{I}$ is the indicator function,

$$\mathbb{I}_A(q) = \begin{cases} 0, & q \notin A \\ 1, & q \in A \end{cases}, q \in Q, A \in \mathcal{B}(Q).$$

In other words, the kernel assigns a probability to a set, $A \in \mathcal{B}(Q)$, by computing the measure of the preimage of that set, $t^{-1}(A)$, averaged over all isomorphisms in Γ. Because each t preserves the target measure, so too will their convolution and, consequently, the Markov transition induced by the kernel.

This construction provides a new perspective on the limited performance of existing algorithms.

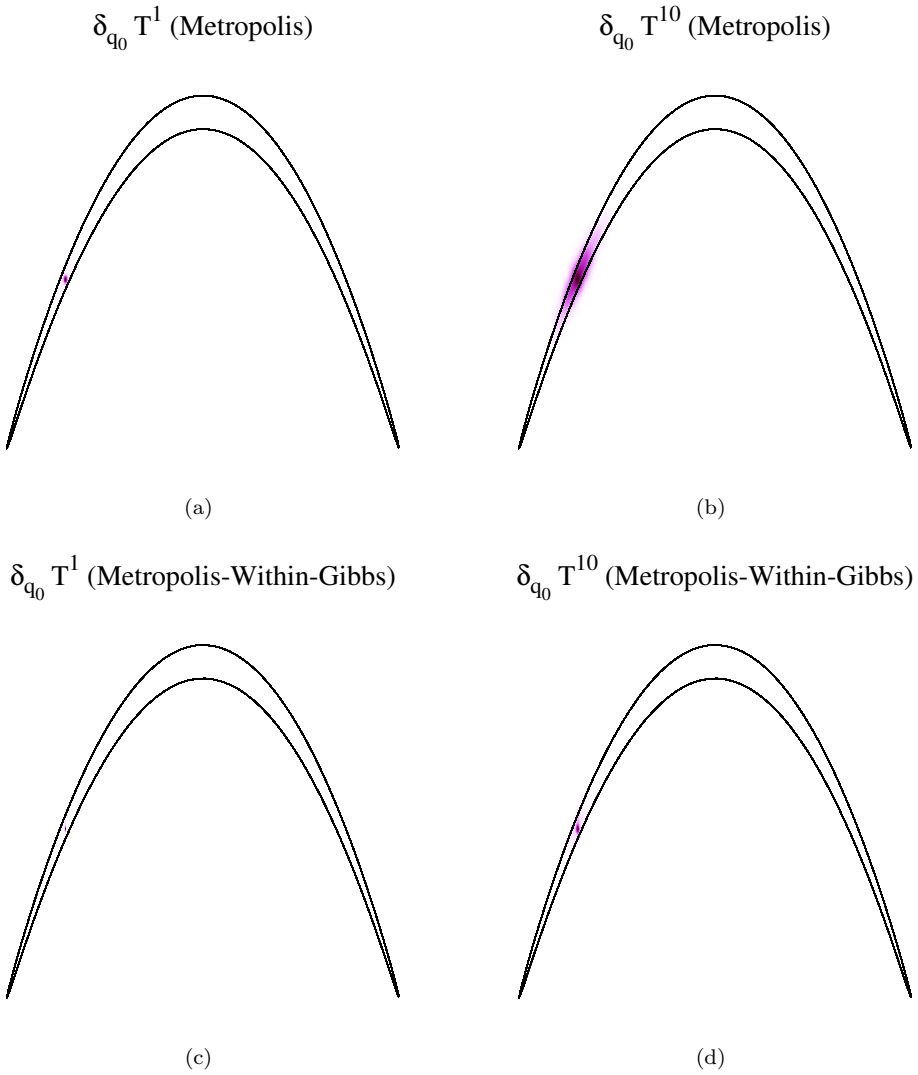


Figure 1. Both (a, b) Random Walk Metropolis and (c, d) the Gibbs sampler are stymied by complex distributions, for example the twisted Gaussian distribution on the sample space $Q = \mathbb{R}^2$ presented in Haario, Saksman and Tamminen (2001), here represented with a 95% probability contour. Even when optimally tuned (Roberts et al., 1997), both Random Walk Metropolis and Random Walk Metropolis-within-Gibbs kernels concentrate around the initial point, even after multiple iterations.

***Example* 1.** *We can consider Gaussian Random Walk Metropolis, for example, as being generated by random, independent translations of each point in the sample space,*

$$\begin{aligned} t_{\epsilon,\eta} : q &\mapsto q + \epsilon \, \mathbb{I}\left(\eta < \frac{f(q + \epsilon)}{f(q)} \right) \\ \epsilon &\sim \mathcal{N}(0, \Sigma) \\ \eta &\sim U[0, 1] \, , \end{aligned}$$

where f is the density of π with respect to the Lebesgue measure on $\mathbb{R}^n$. When targeting complex distributions either ϵ or the support of the indicator will be small and the resulting translations barely perturb the initial state.

***Example* 2.** *The random scan Gibbs sampler is induced by axis-aligned translations,*

$$\begin{aligned} t_{i,\eta} : q_i &\mapsto P_i^{-1}(\eta) \\ i &\sim U\{1, \ldots, n\} \\ \eta &\sim U[0, 1] \, , \end{aligned}$$

where

$$P_i(q_i) = \int_{\infty}^{q_i} \pi(\mathrm{d}\tilde{q}_i | q)$$

is the cumulative distribution function of the ith conditional measure. When the target distribution is strongly correlated, the conditional measures concentrate near the initial q and, as above, the translations are stunted.

In order to define a Markov kernel that remains efficient in difficult problems we need measure-preserving maps whose domains are not limited to local exploration. Realizations of Langevin diffusions (Øksendal, 2003), for example, yield measure-preserving maps that diffuse across the entire target distribution. Unfortunately that diffusion tends to expand across the target measures only slowly (Figure 2): for any finite diffusion time the resulting Langevin kernels are localized around the initial point (Figure 3). What we need are more coherent maps that avoid such diffusive behavior.

One potential candidate for coherent maps are *flows*. A flow, $\{\phi_t\}$, is a family of isomorphisms parameterized by a time, t,

$$\phi_t : Q \to Q, \, \forall t \in \mathbb{R},$$

that form a one-dimensional group on composition,

$$\begin{aligned} \phi_t \circ \phi_s &= \phi_{s+t} \\ \phi_t^{-1} &= \phi_{-t} \\ \phi_0 &= \mathrm{Id}_Q, \end{aligned}$$

where Id_Q is the natural identity map on Q. Because the inverse of a map is given only by negating t, as the time is increased the resulting ϕ_t pushes points away from their initial positions and avoids localized exploration.Our final obstacle is in engineering a flow comprised of measure-preserving maps.

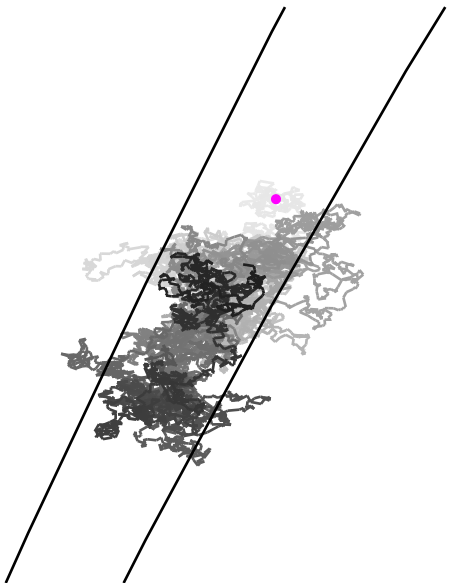


Figure 2. Langevin trajectories are, by construction, diffusive, and are just as likely to double back as they are to move forward. Consequently even as the diffusion time grows, here to $t = 1000$ as the trajectory darkens, realizations of a Langevin diffusion targeting the twisted Gaussian distribution from Haario, Saksman and Tamminen (2001) only slowly wander away from the initial point.

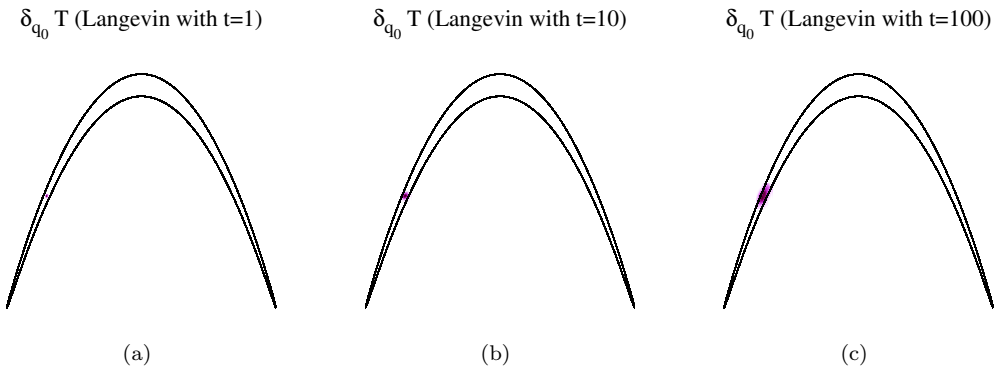


Figure 3. Because of the diffusive nature of the underlying maps, Langevin kernels expand very slowly with increasing diffusion time, t. For any reasonable diffusion time the resulting kernels will concentrate around the initial point, as seen here for a Langevin diffusion targeting the twisted Gaussian distribution from Haario, Saksman and Tamminen (2001).

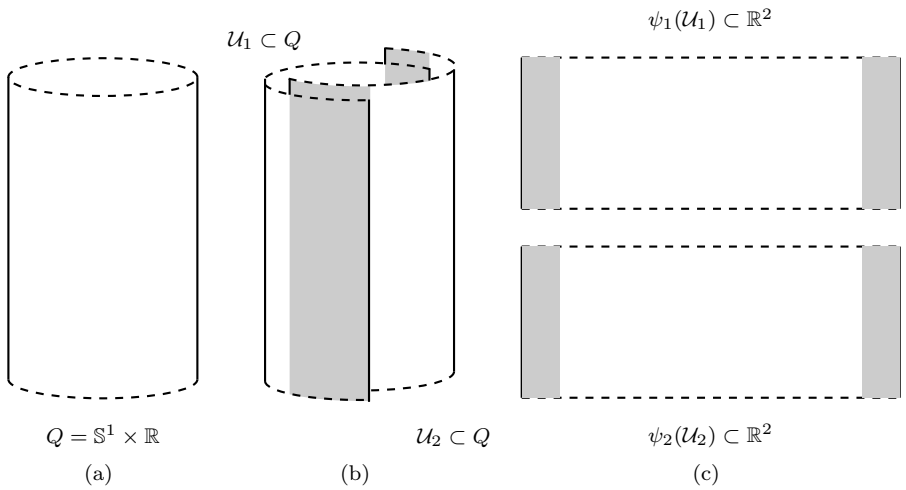


Figure 4. (a) The cylinder, $Q = \mathbb{S}^1 \times \mathbb{R}$, is a nontrivial example of a manifold. Although not globally equivalent to a Euclidean space, (b) the cylinder can be covered in two neighborhoods (c) that are themselves isomorphic to an open neighborhood in $\mathbb{R}^2$. The manifold is smooth when $\psi_1 \circ \psi_2^{-1} : \mathbb{R}^2 \to \mathbb{R}^2$ is a smooth function wherever the two neighborhoods intersect (the intersections here shown in gray).

Flows are particularly natural on the *smooth manifolds* of differential geometry, and flows that preserve a given target measure can be engineered on one special class of smooth manifolds known as *symplectic manifolds*. If we can understand these manifolds probabilistically then we can take advantage of their properties to build Markov kernels with small autocorrelations for even the complex, high-dimensional target distributions of practical interest.

2. Measures on Manifolds

In this section we review probability measures on smooth manifolds of increasing sophistication, culminating in the construction of measure-preserving flows.

Although we will relate each result to probabilistic theory and introduce intuition where we can, the formal details in the following require a working knowledge of differential geometry up to Lee (2013). We will also use the notation therein throughout the paper. For readers new to the subject but interested in learning more, we recommend the introduction in Baez and Muniain (1994), the applications in Schutz (1980); José and Saletan (1998), and then finally Lee (2013). The theory of symplectic geometry in which we will be particularly interested is reviewed in Schutz (1980); José and Saletan (1998); Lee (2013), with Cannas da Silva (2001) providing the most modern and thorough coverage of the subject.

Smooth manifolds generalize the Euclidean space of real numbers and the corresponding calculus; in particular, a smooth manifold need only look *locally* like a Euclidean space (Figures 4). These more general spaces includes Lie groups, Stiefel manifolds, and other spaces becoming common in contemporary applications (Byrne and Girolami, 2013), not to mention Euclidean spaces as a special case. It does not, however, include any manifold with a discrete topology such as tree spaces.

Formally, we assume that our sample space, Q, satisfies the properties of a smooth, connected, and orientable n-dimensional manifold. Specifically we require that Q be a Hausdorff and second-countable topological space that is locally homeomorphic to $\mathbb{R}^n$ and equipped with a differential structure,

$$\{\mathcal{U}_\alpha, \psi_\alpha\}_{\alpha \in I},$$

consisting of open sets in Q,

$$\mathcal{U}_\alpha \subset Q,$$

and homeomorphic charts,

$$\psi_\alpha : \mathcal{U}_\alpha \to \mathcal{V}_\alpha \subset \mathbb{R}^n,$$

that are smooth functions whenever their domains overlap (Figure 4),

$$\psi_\beta \circ \psi_\alpha^{-1} \in C^\infty(\mathbb{R}^n), \forall \alpha, \beta \,|\, \mathcal{U}_\alpha \cap \mathcal{U}_\beta \neq \emptyset. \tag{2}$$

Coordinates subordinate to a chart,

$$\begin{aligned} q^i : \mathcal{U}_\alpha &\to \mathbb{R} \\ q &\mapsto \varpi_i \circ \psi_\alpha(q), \end{aligned}$$

where ϖ_i is the ith Euclidean projection on the image of ψ_α, provide local parameterizations of the manifold convenient for explicit calculations.

This differential structure allows us to define calculus on manifolds by applying concepts from real analysis in each chart. The differential properties of a function $f : Q \to \mathbb{R}$, for example, can be studied by considering the real functions,

$$f \circ \psi_\alpha^{-1} : \mathbb{R}^n \to \mathbb{R};$$

because the charts are smooth in their overlap (2), these local properties define a consistent global definition of smoothness.

Ultimately these properties manifest as geometric objects on Q, most importantly vector fields and differential k-forms. Informally, vector fields specify directions and magnitudes at each point in the manifold while k-forms define multilinear, antisymmetric maps of k such vectors to $\mathbb{R}$. If we consider n linearly-independent vector fields as defining infinitesimal parallelepipeds at every point in space, then the action of n-forms provides a local sense of volume and, consequently, integration. In particular, when the manifold is orientable we can define n-forms that are everywhere positive and a geometric notion of a measure.

Here we consider the probabilistic interpretation of these volume forms, first on smooth manifolds in general and then on smooth manifolds with additional structure: fiber bundles, Riemannian manifolds, and symplectic manifolds. Symplectic manifolds will be particularly important as they naturally provide measure-preserving flows. Proofs of intermediate lemmas are presented in Appendix A.

2.1. Smooth Measures on Generic Smooth Manifolds

Formally, volume forms are defined as positive, top-rank differential forms,

$$\mathcal{M}(Q) \equiv \{\mu \in \Omega^n(Q) \,|\, \mu_q > 0, \forall q \in Q\},$$

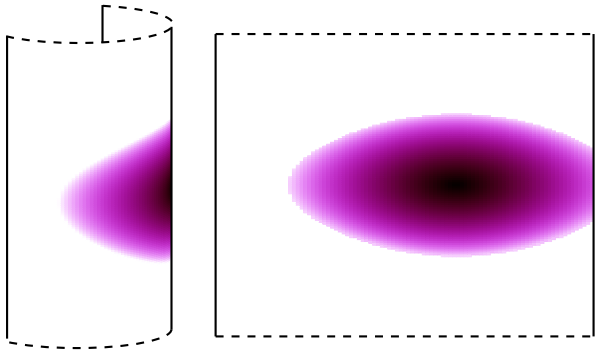


Figure 5. (a) In the neighborhood of a chart, any top-rank differential form is specified by its density, $\mu(q^1, \ldots, q^n)$, with respect to the coordinate volume, $\mu = \mu(q^1, \ldots, q^n)\, \mathrm{d}q^1 \wedge \ldots \wedge \mathrm{d}q^n$, (b) which pushes forward to a density with respect to the Lebesgue measure in the image of the corresponding chart. By smoothly patching together these equivalences, Lemma 1 demonstrates that these forms are in fact measures.

where $\Omega^n(Q)$ is the space of n-forms on Q. By leveraging the local equivalence to Euclidean space, we can show that these volume forms satisfy all of the properties of σ-finite measures on Q (Figure 5).

Lemma 1. *If Q is an oriented, smooth manifold then $\mathcal{M}(Q)$ is non-empty and its elements are σ-finite measures on Q.*

We will refer to elements of $\mathcal{M}(Q)$ as *smooth measures* on Q.

Because of the local compactness of Q, the elements of $\mathcal{M}(Q)$ are not just measures but also Radon measures. As expected from the Riesz Representation Theorem (Folland, 1999), any such element also serves as a linear functional via the usual geometric definition of integration,

$$\begin{aligned} \mu : L^1(Q, \mu) &\to \mathbb{R} \\ f &\mapsto \int_Q f \mu. \end{aligned}$$

Consequently, $(Q, \mathcal{M}(Q))$ is also a Radon space, which guarantees the existence of various probabilistic objects such as disintegrations as discussed below.

Ultimately we are not interested in the whole of $\mathcal{M}(Q)$ but rather $\mathcal{P}(Q)$, the subset of volume forms with unit integral,

$$\mathcal{P}(Q) = \left\{ \pi \in \mathcal{M}(Q) \left| \int_Q \pi = 1 \right. \right\},$$

which serve as probability measures. Because we can always normalize measures, $\mathcal{P}(Q)$ is equivalent to the finite elements of $\mathcal{M}(Q)$,

$$\widetilde{\mathcal{M}}(Q) = \left\{ \pi \in \mathcal{M}(Q) \left| \int_Q \pi < \infty \right. \right\},$$

modulo their normalizations.

Corollary 2. *If Q is an oriented, smooth manifold then $\widetilde{\mathcal{M}}(Q)$, and hence $\mathcal{P}(Q)$, is non-empty.*

Proof. Because the manifold is paracompact, the prototypical measure constructed in Lemma 1 can always be chosen such that the measure of the entire manifold is finite.

□

2.2. Smooth Measures on Fiber Bundles

Although conditional probability measures are ubiquitous in statistical methodology, they are notoriously subtle objects to rigorously construct in theory (Halmos, 1950). Formally, a conditional probability measure appeals to a measurable function between two generic spaces, $F : R \to S$, to define measures on R for some subsets of S along with an abundance of technicalities. It is only when S is endowed with the quotient topology relative to F (Folland, 1999; Lee, 2011) that we can define *regular conditional probability measures* that shed many of the technicalities and align with common intuition. In practice, regular conditional probability measures are most conveniently constructed as *disintegrations* (Chang and Pollard, 1997; Leão Jr, Fragoso and Ruffino, 2004).

Fiber bundles are smooth manifolds endowed with a measurable map to a lower-dimensional submanifold and the quotient topology necessary to admit disintegrations. Consequently, fiber bundles are the natural geometries for defining conditional and marginal probability measures.

2.2.1. Fiber Bundles

A smooth fiber bundle, $\varpi : Z \to Q$, combines an $(n+k)$-dimensional total space, Z, an n-dimensional base space, Q, and a smooth projection, ϖ, that injectively maps, or *submerses*, the total space into the base space. We will refer to a positively-oriented fiber bundle as a fiber bundle in which both the total space and the base space are positively-oriented and the projection operator is orientation-preserving.

Each fiber,

$$Z_q = \varpi^{-1}(q)\,,$$

is itself a k-dimensional manifold isomorphic to a common fiber space, F, and is naturally immersed into the total space,

$$\iota_q : Z_q \hookrightarrow Z,$$

where ι_q is the inclusion map. We will make heavy use of the fact that there exists a trivializing cover of the base space, $\{\mathcal{U}_\alpha\}$, along with subordinate charts and a partition of unity, where the corresponding total space is isomorphic to a trivial product (Figures 6, 7),

$$\varpi^{-1}(\mathcal{U}_\alpha) \approx \mathcal{U}_\alpha \times F.$$

Vector fields on Z are classified by their action under the projection operator. Vertical vector fields, Y_i, lie in the kernel of the projection operator,

$$\varpi_* Y_i = 0,$$

while horizontal vector fields, $\tilde{X}_i$, push forward to the tangent space of the base space,

$$\varpi_* \tilde{X}_i(z) = X_i(\varpi(z)) \in T_{\varpi(z)} Q,$$

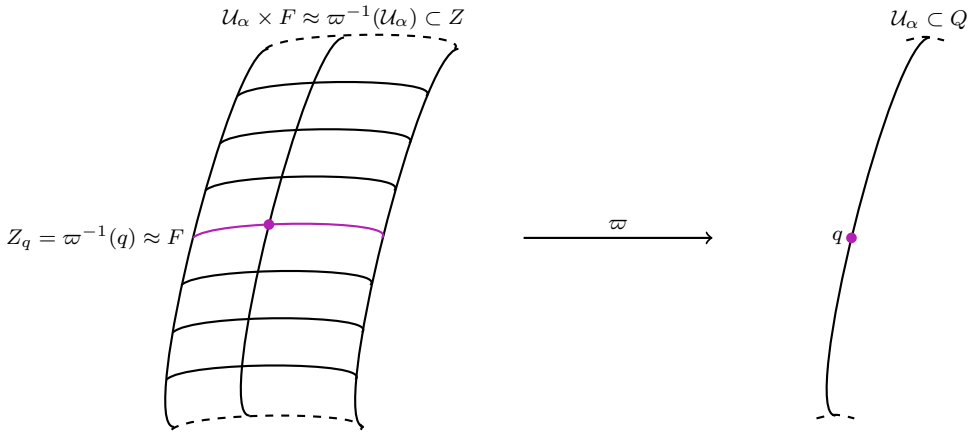


Figure 6. In a local neighborhood, the total space of a fiber bundle, $\varpi^{-1}(\mathcal{U}_\alpha) \subset Z$, is equivalent to attaching a copy of some common fiber space, F, to each point of the base space, $q \in \mathcal{U}_\alpha \subset Q$. Under the projection operator each fiber projects back to the point at which it is attached.

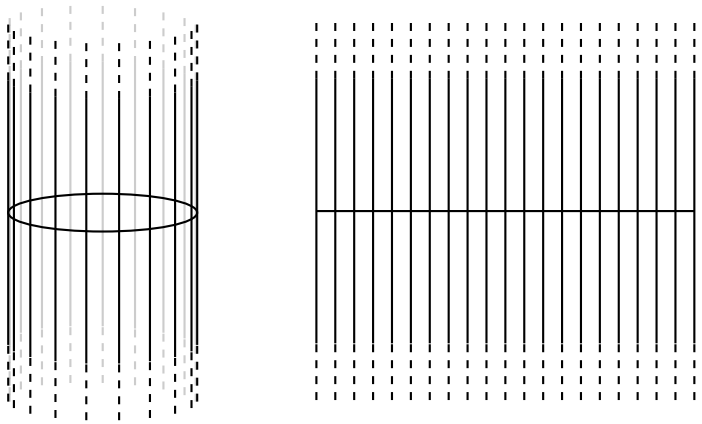


Figure 7. (a) The canonical projection, $\varpi : \mathbb{S}^1 \times \mathbb{R} \to \mathbb{S}^1$, gives the cylinder the structure of a fiber bundle with fiber space $F = \mathbb{R}$. (b) The domain of each chart becomes isomorphic to the product of a neighborhood of the base space, $\mathcal{U}_\alpha \subset \mathbb{S}^1$, and the fiber, $\mathbb{R}$.

where $z \in Z$ and $\varpi(z) \in Q$. Horizontal forms are forms on the total space that vanish when contracted against one or more vertical vector fields.

Note that vector fields on the base space do not uniquely define horizontal vector fields on the total space; a choice of $\tilde{X}_i$ consistent with X_i is called a *horizontal lift* of X_i. More generally we will refer to the *lift* of an object on the base space as the selection of some object on the total space that pushes forward to the corresponding object on the base space.

2.2.2. Disintegrating Fiber Bundles

Because both Z and Q are both smooth manifolds, and hence Radon spaces, the structure of the fiber bundle guarantees the existence of disintegrations with respect to the projection operator (Leão Jr, Fragoso and Ruffino, 2004; Simmons, 2012; Censor and Grandini, 2014) and, under certain regularity conditions, regular conditional probability measures. A substantial benefit of working with smooth manifolds is that we can not only prove the existence of disintegrations but also explicitly construct their geometric equivalents and utilize them in practice.

***Definition* 1.** *Let $(R, \mathcal{B}(R))$ and $(S, \mathcal{B}(S))$ be two measurable spaces with the respective σ-finite measures μ_R and μ_S, and a measurable map, $F : R \to S$, between them. A* disintegration *of μ_R with respect to F and μ_S is a map,*

$$\nu : S \times \mathcal{B}(R) \to \mathbb{R}^+,$$

such that

i $\nu(s, \cdot)$ *is a $\mathcal{B}(R)$-finite measure concentrating on on the level set $F^{-1}(s)$, i.e. for μ_S-almost all s*

$$\nu(s, A) = 0, \, \forall A \in \mathcal{B}(R) \, | A \cap F^{-1}(s) = 0,$$

and for any positive, measurable function $f \in L^1(R, \mu_R)$,

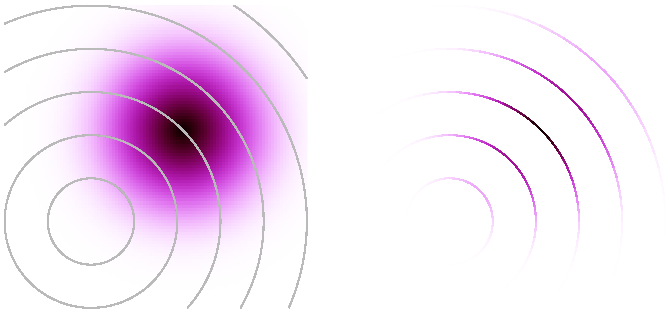
ii $s \mapsto \int_R f(r) \, \nu(s, \mathrm{d}r)$ *is a measurable function for all $s \in S$.*
iii $\int_R f(r) \, \mu_R(\mathrm{d}r) = \int_S \int_{F^{-1}(s)} f(r) \, \nu(s, \mathrm{d}r) \, \mu_S(\mathrm{d}s)$.

In other words, a disintegration is an unnormalized Markov kernel that concentrates on the level sets of F instead of the whole of R (Figure 8). Moreover, if μ_R is finite or F is proper then the pushforward measure,

$$\begin{aligned} \mu_S &= T_* \mu_R \\ \mu_S(B) &= \mu_R\big(F^{-1}(B)\big) \, , \, \forall B \in \mathcal{B}(S) \, , \end{aligned}$$

is σ-finite and known as the marginalization of μ_R with respect to F. In this case the disintegration of μ_R with respect to its pushforward measure becomes a normalized kernel and exactly a regular conditional probability measure. The classic marginalization paradoxes of measure theory (Dawid, Stone and Zidek, 1973) occur when the pushforward of μ_R is not σ-finite and the corresponding disintegration, let alone a regular conditional probability measure, does not exist; we will be careful to explicitly exclude such cases here.

For the smooth manifolds of interest we do not need the full generality of disintegrations, and instead consider the equivalent object restricted to smooth measures.



(a) (b)

Figure 8. The disintegration of (a) a σ-finite measure on the space R with respect to a map $F : R \to S$ and a σ-finite measure on S defines (b) a family of σ-finite measures that concentrate on the level sets of F.

Definition 2. *Let R and S be two smooth, orientable manifolds with the respective smooth measures μ_R and μ_S, and a smooth, orientation-preserving map, $F : R \to S$, between them. A* smooth disintegration *of μ_R with respect to F and μ_S is a map,*

$$\nu : S \times \mathcal{B}(R) \to \mathbb{R}^+,$$

such that

i $\nu(s, \cdot)$ is a smooth measure concentrating on on the level set $F^{-1}(s)$, i.e. for μ_S-almost all s

$$\nu(s, A) = 0, \, \forall A \in \mathcal{B}(R) \, | A \cap F^{-1}(s) = 0,$$

and for any positive, smooth function $g \in L^1(R, \mu_R)$,

ii The function $G(s) = \int_R g(r) \, \nu(s, \mathrm{d}r)$ is integrable with respect to any smooth measure on S.
iii $\int_R g(r) \, \mu_R(\mathrm{d}r) = \int_S \int_{F^{-1}(s)} g(r) \, \nu(s, \mathrm{d}r) \, \mu_S(\mathrm{d}s)$.

Smooth disintegrations have a particularly nice geometric interpretation: Definition 2i implies that these disintegrations define volume forms when pulled back onto the fibers, while Definition 2ii implies that they are smooth objects in the total space (Figure 9). Hence if we want to construct smooth disintegrations geometrically then we should consider the space of k-forms on Z that restrict to finite volume forms on the fibers, i.e. $\omega \in \Omega^k(Z)$ satisfying

$$\iota_q^* \omega > 0$$
$$\int_{Z_q} \iota_q^* \omega < \infty.$$

Note that the finiteness condition is not strictly necessary, but allows us to construct smooth disintegrations independent of the exact measure being disintegrated.

The only subtlety with such a definition is that k-forms on the total space differing by only a horizontal form will restrict to the same volume form on the fibers. Consequently we will consider

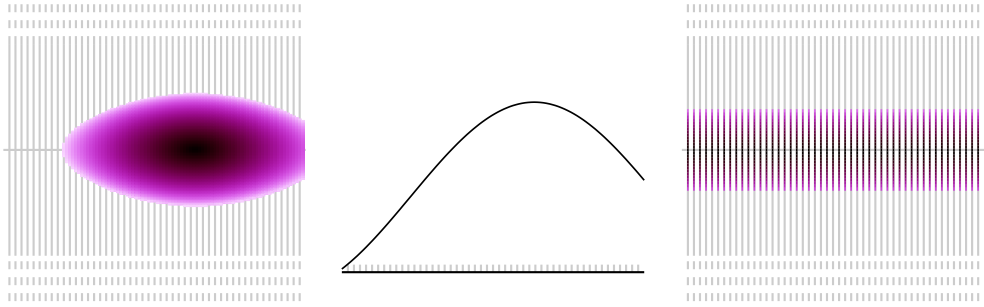


Figure 9. Considering the cylinder as a fiber bundle, $\varpi : \mathbb{S}^1 \times \mathbb{R} \to \mathbb{S}^1$, (a) any joint measure on the total space $\mathbb{S}^1 \times \mathbb{R}$, and (b) any measure on the base space, $\mathbb{S}^1$, define (c) a disintegration that concentrates on the fibers, $\mathbb{R}$. Given any two of these objects we can uniquely construct the third.

the equivalence classes of k-forms up to the addition of horizontal forms,

$$\begin{aligned} \Upsilon(\varpi : Z \to Q) &\equiv \Omega^k(Z) \,/\sim \\ \omega_1 \sim \omega_2 &\Leftrightarrow \omega_1 - \omega_2 \in \Omega^k_H(\varpi : Z \to Q)\,, \end{aligned}$$

where $\Omega^k_H(\varpi : Z \to Q)$ is the space of horizontal k-forms on the total space. As expected from the fact that any smooth manifold is a Radon space, such forms always exist.

Lemma 3. *The space $\Upsilon(\varpi : Z \to Q)$ is convex and nonempty.*

Given a point on the base space, $q \in Q$, the elements of $\Upsilon(\varpi : Z \to Q)$ naturally define smooth measures that concentrate on the fibers.

Lemma 4. *Any element of $\Upsilon(\varpi : Z \to Q)$ defines a smooth measure,*

$$\begin{aligned} \nu : Q \times \mathcal{B}(Z) &\to \mathbb{R}^+ \\ (q, A) &\mapsto \int_{\iota_q(A \cap Z_q)} \iota_q^* \upsilon, \end{aligned}$$

concentrating on the fiber Z_q,

$$\nu(q, A) = 0,\, \forall A \in \mathcal{B}(Z)\,|A \cap Z_q = 0.$$

Finally, any $\upsilon \in \Upsilon(\varpi : Z \to Q)$ satisfies an equivalent of the product rule.

Lemma 5. *Any element $\upsilon \in \Upsilon(\varpi : Z \to Q)$ lifts any smooth measure on the base space, $\mu_Q \in \mathcal{M}(Q)$, to a smooth measure on the total space by*

$$\mu_Z = \varpi^* \mu_Q \wedge \upsilon \in \mathcal{M}(Z)\,.$$

Note the resemblance to the typical measure-theoretic result,

$$\mu_Z(\mathrm{d}z) = \mu_Q(\mathrm{d}q)\, \upsilon(q, \mathrm{d}z).$$

Consequently, the elements of $\Upsilon(\varpi : Z \to Q)$ define smooth disintegrations of any smooth measure on the total space.

Theorem 6. *A positively oriented, smooth fiber bundle admits a smooth disintegration of any smooth measure on the total space, $\mu_Z \in \mathcal{M}(Z)$, with respect to the projection operator and any smooth measure $\mu_Q \in \mathcal{M}(Q)$.*

Proof. From Lemma 5 we know that for any $\upsilon' \in \Upsilon(\varpi : Z \to Q)$ the exterior product $\varpi^* \mu_Q \wedge \upsilon'$ is a smooth measure on the total space, and because the space of smooth measures is one-dimensional we must have

$$\mu_Z = h\, \varpi^* \mu_Q \wedge \upsilon',$$

for some bounded, positive function $h : Z \to \mathbb{R}^+$. Because h is everywhere positive it can be absorbed into υ to define a new, unique element $\upsilon \in \Upsilon(Z)$ such that $\mu_Z = \varpi^* \mu_Q \wedge \upsilon$. With Lemma 3 showing that $\Upsilon(\varpi : Z \to Q)$ is non-empty, such an υ exists for any positively-oriented, smooth fiber bundle.

From Lemma 4, this υ defines a smooth kernel and, hence, satisfies Definition 2i.

If λ_Q is any smooth measure on Q, not necessarily equal to μ_Q, then for any smooth, positive function $g \in L^1(Z, \mu_Z)$ and $G(q) = \int_{Z_q} \iota_q^*(g\, \upsilon)$ we have

$$\int_Q G(q)\, \lambda_Q = \int_Q \left[\int_{Z_q} \iota_q^*(g\, \upsilon) \right] \lambda_Q,$$

or, employing a trivializing cover,

$$\begin{aligned}
\int_Q G(q)\, \lambda_Q &= \sum_\alpha \int_{\mathcal{U}_\alpha} \rho_\alpha \left[\int_{Z_q} \iota_q^*(g_\alpha\, \upsilon) \right] \lambda_Q \\
&= \sum_\alpha \int_{\mathcal{U}_\alpha} \rho_\alpha \left[\int_F \iota_q^*(g_\alpha\, \upsilon) \right] \lambda_Q \\
&= \sum_\alpha \int_{\mathcal{U}_\alpha \times F} \rho_\alpha\, g_\alpha\, \varpi^* \lambda_Q \wedge \upsilon,
\end{aligned}$$

where F here refers to the fiber space. Once again noting that the space of smooth measures on Z is one-dimensional, we must have for some positive, bounded function $h : Z \to \mathbb{R}^+$,

$$\begin{aligned}
\int_Q G(q)\, \lambda_Q &= \sum_\alpha \int_{\mathcal{U}_\alpha \times F} \rho_\alpha\, g_\alpha\, h\, \mu_Z \\
&= \int_Z f\, g\, \mu_Z.
\end{aligned}$$

Because h is bounded the integral is finite given the μ_Z integrability of g, hence υ satisfies Definition 2ii.

Similarly, for any smooth, positive function $g \in L^1(Z, \mu_Z)$,

$$\begin{aligned}
\int_Z g \, \mu_Z &= \int_Z g \, \varpi^* \mu_Q \wedge \upsilon \\
&= \sum_\alpha \int_{\mathcal{U}_\alpha \times F} \rho_\alpha \, g_\alpha \, \varpi^* \mu_Q \wedge \upsilon \\
&= \sum_\alpha \int_{\mathcal{U}_\alpha} \rho_\alpha \left[\int_F \iota_q^*(g_\alpha \upsilon) \right] \mu_Q \\
&= \int_Q \left[\int_{Z_q} \iota_q^*(g \, \upsilon) \right] \mu_Q \\
&= \int_Q \left[\int_{Z_q} g \, \nu(q, \cdot) \right] \mu_Q .
\end{aligned}$$

Because of the finiteness of υ the integral is well-defined and the kernel satisfies Definition 2iii.

Hence for any smooth fiber bundle and smooth measures μ_Z and μ_Q, there exists an $\upsilon \in \Upsilon(\varpi : Z \to Q)$ that induces a smooth disintegration of μ_Z with respect to the projection operator and μ_Q.

□

Ultimately we're not interested in smooth disintegrations but rather regular conditional probability measures. Fortunately, the elements of $\Upsilon(\varpi : Z \to Q)$ are only a normalization away from defining the desired probability measures. To see this, first note that we can immediately define the new space $\Xi(\varpi : Z \to Q) \subset \Upsilon(\varpi : Z \to Q)$ with elements ξ satisfying

$$\begin{aligned}
\iota_q^* \xi &> 0, \\
\int_{Z_q} \iota_q^* \xi &= 1.
\end{aligned}$$

The elements of $\Xi(\varpi : Z \to Q)$ relate a smooth measure on the total space to its pushforward measure with respect to the projection, provided it exists, which is exactly the property needed for the smooth disintegrations to be regular conditional probability measures.

Lemma 7. *Let μ_Z be a smooth measure on the total space of a positively-oriented, smooth fiber bundle with μ_Q the corresponding pushforward measure with respect to the projection operator, $\mu_Q = \varpi_* \mu_Z$. If μ_Q is a smooth measure then $\mu_Z = \varpi^* \mu_Q \wedge \xi$ for a unique element of $\xi \in \Xi(\varpi : Z \to Q)$.*

Consequently the elements of $\Xi(\varpi : Z \to Q)$ also define regular conditional probability measures.

Theorem 8. *Any smooth measure on the total space of a positively-oriented, smooth fiber bundle admits a regular conditional probability measure with respect to the projection operator provided that the pushforward measure with respect to the projection operator is smooth.*

Proof. From Lemma 7 we know that for any smooth measure μ_Z there exists a $\xi \in \Xi(\varpi : Z \to Q)$ such that $\mu_Z = \varpi^* \mu_Q \wedge \xi$ so long as the pushforward measure, μ_Q, is smooth. Applying Theorem 6,

any choice of ξ then defines a smooth disintegration of μ_Z with respect to the projection operator and the pushforward measure and hence the disintegration is a regular conditional probability measure.

□

Although we have shown that elements of $\Xi(\varpi : Z \to Q)$ disintegrate measures on fiber bundles, we have not yet explicitly constructed them. Fortunately the fiber bundle geometry proves productive here as well.

2.2.3. Constructing Smooth Measures From Smooth Disintegrations

The geometric construction of regular conditional probability measures is particularly valuable because it provides an explicit construction for lifting measures on the base space to measures on the total space as well as marginalizing measures on the total space down to the base space.

As shown above, the selection of any element of $\Xi(\varpi : Z \to Q)$ defines a lift of smooth measures on the base space to smooth measures on the total space.

Corollary 9. *If μ_Q is a smooth measure on the base space of a positively-oriented, smooth fiber bundle then for any $\xi \in \Xi(\varpi : Z \to Q)$, $\mu_Z = \varpi^* \mu_Q \wedge \xi$ is a smooth measure on the total space whose pushforward is μ_Q.*

Proof. $\mu_Z = \varpi^* \mu_Q \wedge \xi$ is a smooth measure on the total space by Lemma 5, and Lemma 7 immediately implies that its pushforward is μ_Q.

□

Even before constructing the pushforward measure of a measure on the total space, we can construct its regular conditional probability measure with respect to the projection.

Lemma 10. *Let μ_Z be a smooth measure on the total space of a positively-oriented, smooth fiber bundle whose pushforward measure with respect to the projection operator is smooth, with $U \subset Q$ any neighborhood of the base space that supports a local frame. Within $\varpi^{-1}(U)$, the element $\xi \in \Xi(\varpi : Z \to Q)$*

$$\xi = \frac{\left(\tilde{X}_1, \ldots, \tilde{X}_n\right) \lrcorner \mu_Z}{\mu_Q(X_1, \ldots, X_n)},$$

defines the regular conditional probability measure of μ_Z with respect to the projection operator, where $(X_1, \ldots, X_n)$ is any positively-oriented frame in U satisfying

$$\mu_Q(X_1, \ldots, X_n) < \infty, \forall q \in U$$

and $\left(\tilde{X}_1, \ldots, \tilde{X}_n\right)$ is any corresponding horizontal lift.

The regular conditional probability measure then allows us to validate the geometric construction of the pushforward measure.

Corollary 11. *Let μ_Z be a smooth measure on the total space of a positively-oriented, smooth fiber bundle whose pushforward measure with respect to the projection operator is smooth, with $U \subset Q$ any neighborhood of the base space that supports a local frame. The pushforward measure at any $q \in U$ is given by*

$$\mu_Q(X_1(q), \ldots, X_n(q)) = \int_{Z_q} \iota_q^* \left(\left(\tilde{X}_1, \ldots, \tilde{X}_n \right) \lrcorner \mu_Z \right),$$

where $(X_1, \ldots, X_n)$ is any positively-oriented frame in U satisfying

$$\mu_Q(X_1, \ldots, X_n) < \infty, \forall q \in U$$

and $\left(\tilde{X}_1, \ldots, \tilde{X}_n \right)$ is any corresponding horizontal lift.

Proof. From Lemma 10, the regular conditional probability measure of μ_Z with respect to the projection operator is defined by

$$\xi = \frac{\left(\tilde{X}_1, \ldots, \tilde{X}_n \right) \lrcorner \mu_Z}{\mu_Q(X_1, \ldots, X_n)}.$$

By construction ξ restricts to a unit volume form on any fiber within U, hence

$$\begin{aligned} 1 &= \int_{Z_q} \iota_q^* \xi \\ &= \frac{\int_{Z_q} \iota_q^* \left(\left(\tilde{X}_1, \ldots, \tilde{X}_n \right) \lrcorner \mu_Z \right)}{\mu_Q(X_1(q), \ldots, X_n(q))}, \end{aligned}$$

or

$$\mu_Q(X_1(q), \ldots, X_n(q)) = \int_{Z_q} \iota_q^* \left(\left(\tilde{X}_1, \ldots, \tilde{X}_n \right) \lrcorner \mu_Z \right),$$

as desired.

□

2.3. Measures on Riemannian Manifolds

Once the manifold is endowed with a Riemannian metric, g, the constructions considered above become equivalent to results in classical *geometric measure theory* (Federer, 1969).

In particular, the rigidity of the metric defines projections, and hence regular conditional probability measures, onto any submanifold. The resulting conditional and marginal measures are exactly the co-area and area measures of geometric measure theory.

Moreover, the metric defines a canonical volume form, V_g, on the manifold,

$$V_g = \sqrt{|g|}\, \mathrm{d}q^1 \wedge \ldots \wedge \mathrm{d}q^n.$$

Probabilistically, V_g is a *Hausdorff* measure that generalizes the Lesbegue measure on $\mathbb{R}^n$. If the manifold is globally isomorphic to $\mathbb{R}^n$ then Hausdorff measure reduces to the usual Lebesgue measure.

2.4. Measures on Symplectic Manifolds

A symplectic manifold is an even-dimensional manifold, M, endowed with a non-degenerate symplectic form, $\omega \in \Omega^2(M)$. Unlike Riemannian metrics, there are no local invariants that distinguish between different choices of the symplectic form: within the neighborhood of any chart all symplectic forms are isomorphic to each other and to canonical symplectic form,

$$\omega = \sum_{i=1}^{n} \mathrm{d}q^i \wedge \mathrm{d}p_i,$$

where $\left(q^1, \ldots, q^n, p_1, \ldots, p_n\right)$ are called canonical or Darboux coordinates.

From our perspective, the critical property of symplectic manifolds is that the symplectic form admits not only a canonical family of smooth measures but also a flow that preserves those measures. This structure will be the fundamental basis of Hamiltonian Monte Carlo and hence pivotal to a theoretical understanding of the algorithm.

2.4.1. The Symplectic Measure

Wedging the non-degenerate symplectic form together,

$$\Omega = \bigwedge_{i=1}^{n} \omega,$$

yields a canonical volume form on the manifold.

The equivalence of symplectic forms also ensures that the symplectic volumes, given in local coordinates as

$$\Omega = n! \left(\mathrm{d}q^1 \wedge \ldots \wedge \mathrm{d}q^n \wedge \mathrm{d}p_1 \wedge \ldots \wedge \mathrm{d}p_n\right),$$

are also equivalent locally.

2.4.2. Hamiltonian Systems and Canonical Measures

A symplectic manifold becomes a Hamiltonian system with the selection of a smooth Hamiltonian function,

$$H : M \to \mathbb{R}.$$

Together with the symplectic form, a Hamiltonian defines a corresponding vector field,

$$\mathrm{d}H = \omega(X_H, \cdot)$$

naturally suited to the Hamiltonian system. In particular, the vector field preserves both the symplectic measure and the Hamiltonian,

$$\mathcal{L}_{X_H} \Omega = \mathcal{L}_{X_H} H = 0,$$

where $\mathcal{L}_{X_H}$ is the Lie derivative along the Hamiltonian vector field. Consequently any measure of the form

$$e^{-\beta H} \Omega, \, \beta \in \mathbb{R}^+,$$

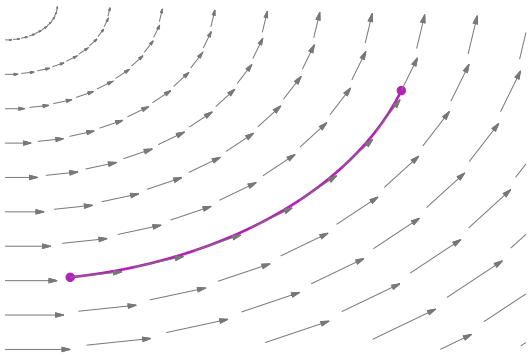


Figure 10. Hamiltonian flow is given by dragging points along the integral curves of the corresponding Hamiltonian vector field. Because they preserve the symplectic measure, Hamiltonian vector fields are said to be *divergenceless* and, consequently, the resulting flow preserves any canonical distribution.

known collectively as *Gibbs measures* or *canonical distributions* (Souriau, 1997), is invariant to the flow generated by the Hamiltonian vector field (Figure 10),

$$\left(\phi_t^H\right)_* \left(e^{-\beta H}\Omega\right) = e^{-\beta H}\Omega,$$

where

$$X_H = \left.\frac{\mathrm{d}\phi_t^H}{\mathrm{d}t}\right|_{t=0}.$$

The level sets of the Hamiltonian,

$$H^{-1}(E) = \{z \in M | H(z) = E\},$$

decompose into *regular level sets* containing only regular points of the Hamiltonian and *critical level sets* which contain at least one critical point of the Hamiltonian. When the critical level sets are removed from the manifold it decomposes into disconnected components, $M = \coprod_i M_i$, each of which foliates into level sets that are diffeomorphic to some common manifold (Figure 11). Consequently each $H : M_i \to \mathbb{R}$ becomes a smooth fiber bundle with the level sets taking the role of the fibers.

Provide that it is finite,

$$\int_M e^{-\beta H}\Omega < \infty,$$

the canonical distribution becomes a probability measure upon normalization,

$$\pi = \frac{e^{-\beta H}\Omega}{\int_M e^{-\beta H}\Omega}.$$

Applying Lemma 10, each component of the excised canonical distribution then disintegrates into *microcanonical distributions* on the level sets,

$$\pi_{H^{-1}(E)} = \frac{v \lrcorner \Omega}{\int_{H^{-1}(E)} \iota_E^* \left(v \lrcorner \Omega\right)}.$$

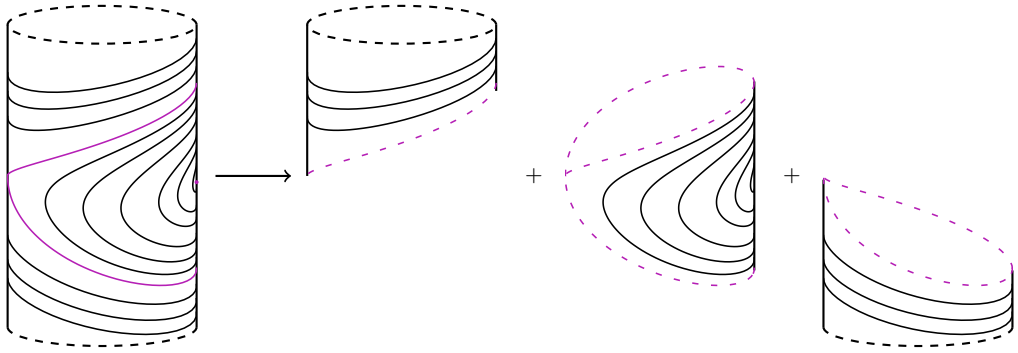


Figure 11. When the cylinder, $\mathbb{S}^1 \times \mathbb{R}$ is endowed with a symplectic structure and Hamiltonian function, the level sets of a Hamiltonian function foliate the manifold. Upon removing the critical level sets, here shown in purple, the cylinder decomposes into three components, each of which becomes a smooth fiber bundle with fiber space $F = \mathbb{S}^1$.

Similarly, the pushforward measure on $\mathbb{R}$ is given by Lemma 11,

$$H_{*}\pi = \frac{e^{-\beta E}}{\int_{M} e^{-\beta H} \Omega} \frac{\left(\int_{H^{-1}(E)} \iota_{E}^{*} \left(v \lrcorner \Omega \right) \right)}{\mathrm{d}H(v)} \mathrm{d}E,$$

where v is any horizontal vector field satisfying $\mathrm{d}H(v) > 0$ to ensure that it is positively-oriented with respect to the image of H. Because the critical level sets have zero measure with respect to the canonical distribution, the disintegration on the excised manifold defines a valid disintegration of original manifold as well. For more on non-geometric constructions of the microcanonical distribution see Draganescu, Lehoucq and Tupper (2009).

The disintegration of the canonical distribution is also compatible with the Hamiltonian flow.

Lemma 12. *Let (M, Ω, H) be a Hamiltonian system with the finite and smooth canonical measure, $\mu = e^{-\beta H} \Omega$. The microcanonical distribution on the level set $H^{-1}(E)$,*

$$\pi_{H^{-1}(E)} = \frac{v \lrcorner \Omega}{\int_{H^{-1}(E)} \iota_{E}^{*} \left(v \lrcorner \Omega \right)},$$

is invariant to the corresponding Hamiltonian flow restricted to the level set, $\phi_{t}^{H} \big|_{H^{-1}(E)}$.

The density of the pushforward of the symplectic measure relative to the Lebesgue measure,

$$d(E) = \frac{\mathrm{d} \left(H_{*} \Omega \right)}{\mathrm{d}E} = \frac{\int_{H^{-1}(E)} \iota_{E}^{*} \left(v \lrcorner \Omega \right)}{\mathrm{d}H(v)},$$

is known as the density of states in the statistical mechanics literature (Kardar, 2007).

3. Hamiltonian Monte Carlo

Although Hamiltonian systems feature exactly the kind of measure-preserving flow that could generate an efficient Markov transition, there is no canonical way of endowing a given probability space with a symplectic form, let alone a Hamiltonian. In order take advantage of Hamiltonian flow we need to consider not the sample space of interest but rather its *cotangent bundle*.

In this section we develop the formal construction of Hamiltonian Monte Carlo and identify how the theory informs practical considerations in both implementation and optimal tuning. Lastly we reconsider a few existing Hamiltonian Monte Carlo implementations with this theory in mind.

3.1. Formal Construction

The key to Hamiltonian Monte Carlo is that the cotangent bundle of the sample space, T^*Q, is endowed with both a canonical fiber bundle structure, $\varpi : T^*Q \to Q$, and a canonical symplectic form. If we can lift the target distribution onto the cotangent bundle then we can construct an appropriate Hamiltonian system and utilize its Hamiltonian flow to generate a powerful Markov kernel. When the sample space is also endowed with a Riemannian metric this construction becomes particularly straightforward.

3.1.1. Constructing a Hamiltonian System

By Corollary 9, the target distribution, π, is lifted onto the cotangent bundle with the choice of a smooth disintegration, $\xi \in \Xi(\varpi : T^*Q \to Q)$,

$$\pi_H = \varpi^* \pi \wedge \xi.$$

Because π_H is a smooth probability measure it must be of the form of a canonical distribution for some Hamiltonian $H : T^*Q \to \mathbb{R}$,

$$\pi_H = e^{-H} \, \Omega,$$

with β taken to be unity without loss of generality. In other words, the choice of a disintegration defines not only a lift onto the cotangent bundle but also a Hamiltonian system (Figure 12) with the Hamiltonian

$$H = -\log \frac{\mathrm{d}\left(\varpi^* \pi \wedge \xi\right)}{\mathrm{d}\Omega}.$$

Although this construction is global, it is often more conveniently implemented in local coordinates. Consider first an open subset of the sample space, $\mathcal{U}_\alpha \subset Q$, in which the target distribution decomposes as

$$\pi = e^{-V} \mathrm{d}q^1 \wedge \ldots \wedge \mathrm{d}q^n.$$

Here e^{-V} is the Radon–Nikodym derivative of the target measure with respect to the pullback of the Lebesgue measure on the image of the local chart. Following the natural analogy to the physical application of Hamiltonian systems, we will refer to V as the *potential energy*.

In the corresponding neighborhood of the cotangent bundle, $\varpi^{-1}(\mathcal{U}_\alpha) \subset T^*Q$, the smooth disintegration, ξ, similarly decomposes into,

$$\xi = e^{-K} \mathrm{d}p_1 \wedge \ldots \wedge \mathrm{d}p_n + \text{horizontal } n\text{-forms}.$$

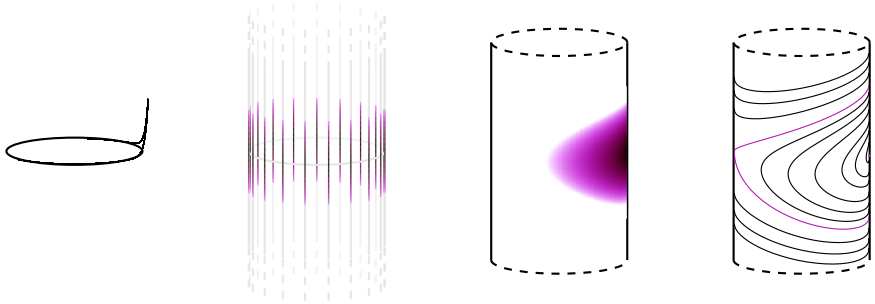


Figure 12. (a) Hamiltonian Monte Carlo begins with a target measure on the base space, for example $Q = \mathbb{S}^1$. (b) The choice of a disintegration on the cotangent bundle, $T^*Q = \mathbb{S}^1 \times \mathbb{R}$, defines (c) a joint measure on the cotangent bundle which immediately defines (d) a Hamiltonian system given the canonical symplectic structure. The Hamiltonian flow of this system is then used to construct an efficient Markov transition.

When ξ is pulled back onto a fiber all of the horizontal n-forms vanish and e^{-K} can be considered the Radon–Nikodym derivative of the disintegration restricted to a fiber with respect to the Lebesgue measure on that fiber. Appealing to the physics conventions once again, we call K as the *kinetic energy*.

Locally the lift onto the cotangent bundle becomes

$$\begin{aligned} \pi_H &= \varpi^* \pi \wedge \pi_q \\ &= e^{-(K+V)} \mathrm{d}q^1 \wedge \ldots \wedge \mathrm{d}q^n \wedge \mathrm{d}p_1 \wedge \ldots \wedge \mathrm{d}p_n \\ &= e^{-H} \Omega, \end{aligned}$$

with the Hamiltonian

$$H = -\log \frac{\mathrm{d}\pi_H}{\mathrm{d}\Omega} = K + V,$$

taking a form familiar from classical mechanics (José and Saletan, 1998).

A particular danger of the local perspective is that neither the potential energy, V, or the kinetic energy, K, are proper scalar functions. Both depend on the choice of chart and introduce a log determinant of the Jacobian when transitioning between charts with coordinates q and q',

$$\begin{aligned} V &\mapsto V + \log \left| \frac{\partial q}{\partial q'} \right| \\ K &\mapsto K - \log \left| \frac{\partial q}{\partial q'} \right|; \end{aligned}$$

only when V and K are summed do the chart-dependent terms cancel to give a scalar Hamiltonian. When these local terms are used to implement Hamiltonian Monte Carlo care must be taken to avoid any sensitivity to the arbitrary choice of chart, which usually manifests as pathological behavior in the algorithm.

3.1.2. Constructing a Markov Transition

The Hamiltonian flow on the cotangent bundle generates isomorphisms that preserve π_H, but in order to define an isomorphism on the sample space we first need to map to the cotangent bundle and back.

If q were drawn from the target measure then we could generate an exact sample from π_H by sampling directly from the measure on the corresponding fiber,

$$p \sim \iota_q^* \xi = e^{-K} \mathrm{d}p_1 \wedge \ldots \wedge \mathrm{d}p_n.$$

In other words, sampling along the fiber defines a lift from π to π_H,

$$\begin{aligned} \lambda : Q &\to T^*Q \\ q &\mapsto (p, q), p \sim \iota_q^* \xi. \end{aligned}$$

In order to return to the sample space we use the canonical projection, which by construction maps π_H back into its pushforward, π.

Together we have a random lift,

$$\begin{aligned} \lambda : Q &\to T^*Q \\ \lambda_* \pi &= \pi_H, \end{aligned}$$

the Hamiltonian flow,

$$\begin{aligned} \phi_t^H : T^*Q &\to T^*Q \\ \left(\phi_t^H\right)_* \pi_H &= \pi_H, \end{aligned}$$

and finally the projection,

$$\begin{aligned} \varpi : T^*Q &\to Q \\ \varpi_* \pi_H &= \pi. \end{aligned}$$

Composing the maps together,

$$\phi_{\mathrm{HMC}} = \varpi \circ \phi_t^H \circ \lambda$$

yields exactly the desired measure-preserving isomorphism,

$$\begin{aligned} \phi_{\mathrm{HMC}} : Q &\to Q \\ \left(\phi_{\mathrm{HMC}}\right)_* \pi &= \pi, \end{aligned}$$

for which we have been looking.

Finally, the measure on λ, and possibly a measure on the integration time, t, specifies a measure on ϕ_{HMC} from which we can define a Hamiltonian kernel via (1).

3.1.3. Constructing an Explicit Disintegration

The only obstacle with implementing Hamiltonian Monte Carlo as constructed is that the disintegration is left completely unspecified. Outside of needing to sample from $\iota_q^* \xi$ there is little motivation for an explicit choice.

This choice of a smooth disintegration, however, is greatly facilitated by endowing the base manifold with a Riemannian metric, g, and its inverse, g^{-1}, which provide two canonical objects from which we can construct a kinetic energy and, consequently, a disintegration. If we denote by $\tilde{p}(z)$ the element of $T^*_{\varpi(z)}Q$ identified by $z \in T^*Q$, then the metric immediately defines a quadratic form, $g^{-1}(\tilde{p}(z), \tilde{p}(z))$ and a density, $|g(\varpi(z))|$. From the perspective of geometry measure theory this latter term is just the Hausdorff density; in molecular dynamics it is known as the Fixman potential (Fixman, 1978).

Noting that the quadratic function is a scalar function whereas the log density transforms like the kinetic energy, an immediate candidate for the kinetic energy is given by simply summing the two together,

$$K(z) = \frac{1}{2} g^{-1}(\tilde{p}(z), \tilde{p}(z)) + \frac{1}{2} \log |g(\varpi(z))| + \text{const},$$

or in coordinates,

$$K(p, q) = \frac{1}{2} \sum_{i,j=1}^{n} p_i p_j \left(g^{-1}(q)\right)^{ij} + \frac{1}{2} \log |g(q)| + \text{const},$$

which defines a Gaussian distribution on the fibers,

$$\iota_q^* \xi = e^{-K} \mathrm{d}p_1 \wedge \ldots \wedge \mathrm{d}p_n = \mathcal{N}(0, g).$$

Using these same two ingredients we could also construct, for example, a multivariate Student's t distribution,

$$\begin{aligned} K(z) &= \frac{\nu + n}{2} \log \left(1 + \frac{1}{\nu} g^{-1}(\tilde{p}(z), \tilde{p}(z))\right) + \frac{1}{2} \log |g(\varpi(z))| + \text{const} \\ \iota_q^* \xi &= t_\nu(0, g), \end{aligned}$$

or any distribution whose sufficient statistic is the Mahalanobis distance.

When g is taken to be Euclidean the resulting algorithm is exactly the Hamiltonian Monte Carlo implementation that has dominated both the literature and applications to date; we refer to this implementation as *Euclidean Hamiltonian Monte Carlo*. The more general case, where g varies with position, is exactly *Riemannian Hamiltonian Monte Carlo* (Girolami and Calderhead, 2011) which has shown promising success when the metric is used to correct for the nonlinearities of the target distribution. In both cases, the natural geometric motivation for the choice of disintegration helps to explain why the resulting algorithms have proven so successful in practice.

3.2. Practical Implementation

Ultimately the Hamiltonian Monte Carlo transition constructed above is a only a mathematical abstraction until we are able to simulate the Hamiltonian flow by solving a system of highly-nonlinear, coupled ordinary differential equations. At this stage the algorithm is vulnerable to a host of pathologies and we have to heed the theory carefully.

The numerical solution of Hamiltonian flow is a well-researched subject and many efficient integrators are available. Of particular importance are symplectic integrators which utilize the underlying symplectic geometry to exactly preserve the symplectic measure with only a small error in the Hamiltonian (Leimkuhler and Reich, 2004; Hairer, Lubich and Wanner, 2006). Because they

preserve the symplectic measure exactly, these integrators are highly accurate even over long integration times.

Formally, symplectic integrators approximate the Hamiltonian flow by composing the flows generated from individual terms in the Hamiltonian. For example, one second-order symplectic integrator approximating the flow from the Hamiltonian $H = H_1 + H_2$ is given by

$$\phi^H_{\delta t} = \phi^{H_1}_{\delta t/2} \circ \phi^{H_2}_{\delta t} \circ \phi^{H_1}_{\delta t/2} + \mathcal{O}\left(\delta t^2\right).$$

The choice of each component, H_i, and the integration of the resulting flow requires particular care. If the flows are not solved exactly then the resulting integrator no longer preserves the symplectic measure and the accuracy plummets. Moreover, each component must be a scalar function on the cotangent bundle: although one might be tempted to take $H_1 = V$ and $H_2 = K$, for example, this would not yield a symplectic integrator as V and K are not proper scalar functions as discussed in Section 3.1.1. When using a Gaussian kinetic energy as described above, a proper decomposition is given by

$$\begin{aligned} H_1 &= \frac{1}{2} \log |g(\varpi(z))| + V(\varpi(z)) \\ H_2 &= \frac{1}{2} g^{-1}(\tilde{p}(z), \tilde{p}(z)). \end{aligned}$$

Although symplectic integrators introduce only small and well-understood errors, those errors will ultimately bias the resulting Markov chain. In order to remove this bias we can consider the Hamiltonian flow not as a transition but rather as a Metropolis proposal on the cotangent bundle and let the acceptance procedure cancel any numerical bias. Because symplectic integrators remain accurate even for high-dimensional systems, their use here is crucial lest the Metropolis acceptance probability fall towards zero.

The only complication with a Metropolis strategy is that the numerical flow, $\Phi^H_{\epsilon,t}$, defines a valid Metropolis proposal only when it is an involution (Tierney, 1998). This can be accomplished by making the measure on the integration time symmetric about 0, or by composing the flow with any operator, R, satisfying

$$\Phi^H_{\epsilon,t} \circ R \circ \Phi^H_{\epsilon,t} = \mathrm{Id}_{T*Q}.$$

For all of the kinetic energies considered above, this is readily accomplished with a parity inversion given in canonical coordinates by

$$R(q, p) = (q, -p).$$

In either case the acceptance probability reduces to

$$a\left(z, R \circ \Phi^H_{\epsilon,t} z\right) = \min\left[1, \exp\left(H\left(R \circ \Phi^H_{\epsilon,t} z\right) - H(z)\right)\right],$$

where $\Phi^H_{\epsilon,t}$ is the symplectic integrator with step size, ϵ, and $z \in T^*Q$.

3.3. Tuning

Although the selection of a disintegration and a symplectic integrator formally define a full implementation of Hamiltonian Monte Carlo, there are still free parameters left unspecified to which

the performance of the implementation will be highly sensitive. In particular, we must set the integration time of the flow, the Riemannian metric, and symplectic integrator step size. All of the machinery developed in our theoretical construction proves essential here, as well.

The Hamiltonian flow generated from a single point may explore the entirety of the corresponding level set or be restricted to a smaller submanifold of the level set, but in either case the trajectory nearly closes in some possibly-long but finite recurrence time, $\tau_{H^{-1}(E)}$ (Petersen, 1989; Zaslavsky, 2005). Taking

$$t \sim U\left(0, \tau_{H^{-1}(E)}\right),$$

would avoid redundant exploration but unfortunately the recurrence time for a given level set is rarely calculable in practice and we must instead resort to approximations. When using a Riemannian geometry, for example, we can appeal to the No-U-Turn sampler (Hoffman and Gelman, 2014; Betancourt, 2013a) which has proven an empirical success.

When using such a Riemannian geometry, however, we must address the fact that the choice of metric is itself a free parameter. One possible criterion to consider is the interaction of the geometry with the symplectic integrator – in the case of a Gaussian kinetic energy, integrators are locally optimized when the metric approximates the Hessian of the potential energy, essentially canceling the local nonlinearities of the target distribution. This motivates using the global covariance of the target distribution for Euclidean Hamiltonian Monte Carlo and the SoftAbs metric (Betancourt, 2013b) for Riemannian Hamiltonian Monte Carlo; for further discussion see Livingstone and Girolami (2014). Although such choices work well in practice, more formal ergodicity considerations are required to define a more rigorous optimality condition.

Lastly we must consider the step size, ϵ, of the symplectic integrator. As the step size is made smaller the integrator will become more accurate but also more expensive – larger step sizes yield cheaper integrators but at the cost of more Metropolis rejections. When the target distribution decomposes into a product of many independent and identically distributed measures, the optimal compromise between these extremes can be computed directly (Beskos et al., 2013). More general constraints on the optimal step size, however, requires a deeper understanding of the interaction between the geometry of the exact Hamiltonian flow and that of the symplectic integrator developed in backwards error analysis (Leimkuhler and Reich, 2004; Hairer, Lubich and Wanner, 2006; Izaguirre and Hampton, 2004). In particular, the microcanonical distribution constructed in Section 2.4.2 plays a crucial role.

3.4. Retrospective Analysis of Existing Work

In addition to providing a framework for developing robust Hamiltonian Monte Carlo methodologies, the formal theory also provides insight into the performance of recently published implementations.

We can, for example, now develop of deeper understanding of the poor scaling of the explicit Lagrangian Dynamical Monte Carlo algorithm (Lan et al., 2012). Here the authors were concerned with the computation burden inherent to the implicit symplectic integrators necessary for Riemannian Hamiltonian Monte Carlo, and introduced an approximate integrator that sacrificed exact symplecticness for explicit updates. As we saw in Section 3.2, however, exact symplecticness is critical to maintaining the exploratory power of an approximate Hamiltonian flow, especially as the dimension of the target distribution increases and numerical errors amplify. Indeed, the empirical results in the paper show that the performance of the approximate integrator suffers with increasing dimension of the target distribution. The formal theory enables an understanding of the compromises, and corresponding vulnerabilities, of such approximations.

Moreover, the import of the Hamiltonian flow elevates the integration time as a fundamental parameter, with the integrator step size accompanying the use of an approximate flow. A common error in empirical optimizations is to reparameterize the integration time and step size as the number of integrator steps, which can obfuscate the optimal settings. For example, Wang, Mohamed and de Freitas (2013) use Bayesian optimization methods to derive an adaptive Hamiltonian Monte Carlo implementation, but they optimize the integrator step size and the number of integrator steps over only a narrow range of values. This leads not only to a narrow range of short integration times that limits the efficacy of the Hamiltonian flow, but also a *step size-dependent* range of integration times that skew the optimization values. Empirical optimizations of Hamiltonian Monte Carlo are most productive when studying the fundamental parameters directly.

In general, care must be taken to not to limit the range of integration times considered lest the performance of the algorithm be misunderstood. For example, restricting the Hamiltonian transitions to only a small fraction of the recurrence time forfeits the efficacy of the flow's coherent exploration. Under this artificial limitation, partial momentum refreshment schemes (Horowitz, 1991; Sohl-Dickstein, Mudigonda and DeWeese, 2014), which compensate for the premature termination of the flow by correlating adjacent transitions, do demonstrate some empirical success. As the restriction is withdrawn and the integration times expand towards the recurrence time, however, the success of such schemes fade. Ultimately, removing such limitations in the first place results in more effective transitions.

4. Future Directions

By appealing to the geometry of Hamiltonian flow we have developed a formal, foundational construction of Hamiltonian Monte Carlo, motivating various implementation details and identifying the properties critical for a high performance algorithm. Indeed, these lessons have already proven critical in the development of high-performance software like Stan (Stan Development Team, 2014). Moving forward, the geometric framework not only admits further understanding and optimization of the algorithm but also suggests connections to other fields and motivates generalizations amenable to an even broader class of target distributions.

4.1. Robust Implementations of Hamiltonian Monte Carlo

Although we have constructed a theoretical framework in which we can pose rigorous optimization criteria for the the integration time, Riemannian metric, and integrator step size, there is much to be done in actually developing and then implementing those criteria. Understanding the ergodicity of Hamiltonian Monte Carlo is a critical step towards this goal, but a daunting technical challenge.

Continued application of both the symplectic and Riemannian geometry underlying implementations of the algorithm will be crucial to constructing a strong formal understanding of the ergodicity of Hamiltonian Monte Carlo and its consequences. Initial applications of metric space methods (Ollivier, 2009; Joulin and Ollivier, 2010), for example, have shown promise (Holmes, Rubinstein-Salzedo and Seiler, 2014), although many technical obstacles, such as the limitations of geodesic completeness, remain.

4.2. Relating Hamiltonian Monte Carlo to Other Fields

The application of tools from differential geometry to statistical problems has rewarded us with a high-performance and robust algorithm. Continuing to synergize seemingly disparate fields of applied mathematics may also prove fruitful in the future.

One evident association is to *molecular dynamics* (Haile, 1992; Frenkel and Smit, 2001; Marx and Hutter, 2009), which tackles expectations of chemical systems with natural Hamiltonian structures. Although care must be taken with the different construction and interpretation of the Hamiltonian from the statistical and the molecular dynamical perspectives, once a Hamiltonian system has been defined the resulting algorithms are identical. Consequently molecular dynamics implementations may provide insight towards improving Hamiltonian Monte Carlo and vice versa.

Additionally, the composition of Hamiltonian flow with a random lift from the sample space onto its cotangent bundle can be considered a second-order stochastic process (Burrage, Lenane and Lythe, 2007; Polettini, 2013), and the theory of these processes has the potential to be a powerful tool in understanding the ergodicity of the algorithm.

Similarly, the ergodicity of Hamiltonian systems has fueled a wealth of research into dynamical systems in the past few decades (Petersen, 1989; Zaslavsky, 2005). The deep geometric results emerging from this field complement those of the statistical theory of Markov chains and provide another perspective on the ultimate performance of Hamiltonian Monte Carlo.

The deterministic flow that powers Hamiltonian Monte Carlo is also reminiscent of various strategies of removing the randomness in Monte Carlo estimation (Caflisch, 1998; Murray and Elliott, 2012; Neal, 2012). The generality of the Hamiltonian construction may provide insight into the optimal compromise between random and deterministic algorithms.

Finally there is the possibility that the theory of measures on manifolds may be of use to the statistical theory of smooth measures in general. The application of differential geometry to Frequentist methods that has consolidated into Information Geometry (Amari and Nagaoka, 2007) has certainly been a great success, and the use Bayesian methods developed here suggests that geometry's domain of applicability may be even broader. As demonstrated above, for example, the geometry of fiber bundles provides a natural setting for the study and implementation of conditional probability measures, generalizing the pioneering work of Tjur (1980).

4.3. Generalizing Hamiltonian Monte Carlo

Although we have made extensive use of geometry in the construction of Hamiltonian Monte Carlo, we have not yet exhausted its utility towards Markov Chain Monte Carlo. In particular, further geometrical considerations suggest tools for targeting multimodal, trans-dimensional, infinite-dimensional, and possibly discrete distributions.

Like most Markov Chain Monte Carlo algorithms, Hamiltonian Monte Carlo has trouble exploring the isolated concentrations of probability inherent to multimodal target distributions. Leveraging the geometry of contact manifolds (Lee, 2013), however, admits not just transitions within a single canonical distribution but also transitions between different canonical distributions. The resulting Adiabatic Monte Carlo provides a geometric parallel to simulated annealing and simulated tempering without being burdened by their common pathologies (Betancourt, 2014).

Trans-dimensional target distributions are another obstacle for Hamiltonian Monte Carlo because of the discrete nature of the model space. Differential geometry may too prove fruitful here with the

Poisson geometries that generalize symplectic geometry by allowing for a symplectic form whose rank need not be constant (Weinstein, 1983).

Many of the properties of smooth manifolds critical to the construction of Hamiltonian Monte Carlo do not immediately extend to the infinite-dimensional target distributions common to functional analysis, such as the study of partial differential equations (Cotter et al., 2013). Algorithms on infinite-dimensional spaces motivated by Hamiltonian Monte Carlo, however, have shown promise (Beskos et al., 2011) and suggest that infinite-dimensional manifolds admit symplectic structures, or the appropriate generalizations thereof.

Finally there is the question of fully discrete spaces from which we cannot apply the theory of smooth manifolds, let alone Hamiltonian systems. Given that Hamiltonian flow can also be thought of as an orbit of the symplectic group, however, there may be more general group-theoretic constructions of measure-preserving orbits that can be applied to discrete spaces.

Acknowledgements

We thank Tom LaGatta for thoughtful comments and discussion on disintegrations and Chris Wendl for invaluable assistance with formal details of the geometric constructions, but claim all errors as our own. The preparation of this paper benefited substantially from the careful readings and recommendations of Saul Jacka, Pierre Jacob, Matt Johnson, Ioannis Kosmidis, Paul Marriott, Yvo Pokern, Sebastian Reich, Daniel Roy, Sebastian Vollmer, and two reviewers. This work was motivated by the initial investigations in Betancourt and Stein (2011).

Michael Betancourt is supported under EPSRC grant EP/J016934/1, Simon Byrne is a EPSRC Postdoctoral Research Fellow under grant EP/K005723/1, Samuel Livingstone is funded by a PhD scholarship from Xerox Research Center Europe, and Mark Girolami is an EPSRC Established Career Research Fellow under grant EP/J016934/1.

Appendix A: Proofs

Here we collect the proofs of the Lemmas introduced in Section 2.

Lemma 1. *If Q is an oriented, smooth manifold then $\mathcal{M}(Q)$ is non-empty and its elements are σ-finite measures on Q.*

Proof. We begin by constructing a prototypical element of $\mathcal{M}(Q)$. In a local chart $\{\mathcal{U}_\alpha, \psi_\alpha\}$ we can construct a positive μ_α as $\mu_\alpha = f_\alpha \, \mathrm{d}q^1 \wedge \ldots \wedge \mathrm{d}q^n$ for any $f_\alpha : \mathcal{U}_\alpha \to \mathbb{R}^+$. Given the positive orientation of Q, the set of μ_α is convex and we can define a global $\mu \in \mathcal{M}(Q)$ by employing a partition of unity subordinate to the $\mathcal{U}_\alpha$,

$$\mu = \sum_\alpha \rho_\alpha \mu_\alpha.$$

To show that any $\mu \in \mathcal{M}(Q)$ is a measure, consider the integral of μ over any $A \in \mathcal{B}(Q)$. By construction

$$\mu(A) = \int_A \mu > 0,$$

leaving us to show that $\mu(A)$ satisfies countable additivity and vanishes when $A = \emptyset$. We proceed by covering A in charts and employing a partition of unity to give

$$\begin{aligned}
\int_A \mu &= \sum_\alpha \int_{A \cap \mathcal{U}_\alpha} \rho_\alpha \, \mu_\alpha \\
&= \sum_\alpha \int_{A \cap \mathcal{U}_\alpha} \rho_\alpha \, f_\alpha \, \mathrm{d}q^1 \wedge \ldots \wedge \mathrm{d}q^n \\
&= \sum_\alpha \int_{\psi_\alpha(A \cap \mathcal{U}_\alpha)} \left(\rho_\alpha f_\alpha \circ \psi_\alpha^{-1} \right) \mathrm{d}^n q,
\end{aligned}$$

where f_α is defined as above and $\mathrm{d}^n q$ is the Lebesgue measure on the image of the charts.

Now each domain of integration is in the σ-algebra of the sample space,

$$A \cap \mathcal{U}_\alpha \in \mathcal{B}(Q) \, ,$$

and, because the charts are diffeomorphisms and hence Lebesgue measurable functions, we must have

$$\psi_\alpha(A \cap \mathcal{U}_\alpha) \in \mathcal{B}(\mathbb{R}^n) \, .$$

Consequently the action of $\mu(A)$ decomposes into a countable number of Lebesgue integrals, and μ immediately inherits countable additivity.

Moreover, $\psi_\alpha \left(\emptyset \cap \mathcal{U}_\alpha \right) = \psi_\alpha \left(\emptyset \right) = \emptyset$ so that, by the same construction as above,

$$\begin{aligned}
\mu(\emptyset) &= \int_\emptyset \mu \\
&= \sum_\alpha \int_{\psi_\alpha(\emptyset \cap \mathcal{U}_\alpha)} \left(\rho_\alpha f_\alpha \circ \psi_\alpha^{-1} \right) \mathrm{d}^n q \\
&= \sum_\alpha \int_\emptyset \left(\rho_\alpha f_\alpha \circ \psi_\alpha^{-1} \right) \mathrm{d}^n q \\
&= 0.
\end{aligned}$$

Finally, because Q is paracompact any $A \in \mathcal{B}(Q)$ admits a locally-finite refinement and, because any $\mu \in \mathcal{M}(Q)$ is smooth, the integral of μ over the elements of any such refinement are also finite. Hence μ itself is σ-finite.

□

Lemma 3. *The space $\Upsilon(\varpi : Z \to Q)$ is convex and nonempty.*

Proof. The convexity of $\Upsilon(\varpi : Z \to Q)$ follows immediately from the convexity of the positivity constraint and admits the construction of elements with a partition of unity.

In any neighborhood of a trivializing cover, $\{\mathcal{U}_\alpha\}$, we have

$$\Upsilon\left(\varpi^{-1}(\mathcal{U}_\alpha)\right) = \mathcal{M}(F)$$

which is nonempty by Corollary 2. Selecting some $\upsilon_\alpha \in \mathcal{M}(F)$ for each α and summing over each neighborhood gives

$$\upsilon = \sum_\alpha \left(\rho_\alpha \circ \varpi \right) \upsilon_\alpha \in \Upsilon(Z) \, .$$

as desired.

□

Lemma 4. *Any element of* $\Upsilon(\varpi : Z \to Q)$ *defines a smooth measure,*

$$\begin{aligned} \nu : Q \times \mathcal{B}(Z) &\to \mathbb{R}^+ \\ (q, A) &\mapsto \int_{\iota_q(A \cap Z_q)} \iota_q^* \upsilon, \end{aligned}$$

concentrating on the fiber Z_q*,*

$$\nu(q, A) = 0, \, \forall A \in \mathcal{B}(Z) \, | A \cap Z_q = 0.$$

Proof. By construction the measure of any $A \in \mathcal{B}(Z)$ is limited to its intersection with the fiber Z_q, concentrating the measure onto the fiber. Moreover, because the immersion, ι_q, preserves the smoothness of υ, $\iota_q^* \upsilon$ is smooth for all $q \in Q$ and the measure must be $\mathcal{B}(F)$-finite. Consequently, the kernel is $\mathcal{B}(Z)$-finite.

□

Lemma 5. *Any element* $\upsilon \in \Upsilon(\varpi : Z \to Q)$ *lifts any smooth measure on the base space,* $\mu_Q \in \mathcal{M}(Q)$*, to a smooth measure on the total space by*

$$\mu_Z = \varpi^* \mu_Q \wedge \upsilon \in \mathcal{M}(Z) \, .$$

Proof. Let $(X_1(q), \ldots, X_n(q))$ be a basis of $T_q Q, q \in Q$, positively-oriented with respect to the μ_Q,

$$\mu_Q \left(X_1(q), \ldots, X_n(q) \right) > 0,$$

and $(Y_1(p), \ldots, Y_k(p))$ a basis of $T_p Z_q, p \in Z_q$, positively-oriented with respect to the pull-back of υ,

$$\iota_q^* \upsilon (Y_1(p), \ldots, Y_k(p)) > 0.$$

Identifying $T_p Z_q$ as a subset of $T_p Z$, any horizontal lift of the $X_i(q)$ to $\tilde{X}_i(q) \in T_{p,q} Z$ yields a positively-oriented basis of the total space, $\left(\tilde{X}_1(q), \ldots, \tilde{X}_n(q), Y_1(p), \ldots, Y_k(p) \right)$.

Now consider the contraction of this positively-oriented basis against $\mu_Z = \varpi^* \mu_Q \wedge \omega$ for any $\omega \in \Omega^k(Z)$. Noting that, by construction, the Y_i are vertical vectors and vanish when contracted against $\varpi^* \mu_Q$, we must have

$$\begin{aligned} \mu_Z \left(\tilde{X}_1(q), \ldots, \tilde{X}_n(q), Y_1(p), \ldots, Y_k(p) \right) &= \varpi^* \mu_Q \wedge \omega \left(\tilde{X}_1(q), \ldots, \tilde{X}_n(q), Y_1(p), \ldots, Y_k(p) \right) \\ &= \varpi^* \mu_Q \left(\tilde{X}_1(q), \ldots, \tilde{X}_n(q) \right) \omega(Y_1(p), \ldots, Y_k(p)) \\ &= \mu_Q(X_1(q), \ldots, X_n(q)) \, \omega(Y_1(p), \ldots, Y_k(p)) \\ &> 0. \end{aligned}$$

Hence μ_Z is a volume form and belongs to $\mathcal{M}(Z)$.

Moreover, adding a horizontal k-form, η, to ω yields the same lift,

$$\begin{aligned}
\mu'_Z\left(\tilde{X}_1(q), \ldots, \tilde{X}_n(q), Y_1(p), \ldots, Y_k(p)\right) & \\
&= \varpi^* \mu_Q \wedge (\omega + \eta)\left(\tilde{X}_1(q), \ldots, \tilde{X}_n(q), Y_1(p), \ldots, Y_k(p)\right) \\
&= \varpi^* \mu_Q\left(\tilde{X}_1(q), \ldots, \tilde{X}_n(q)\right) \omega(Y_1(p), \ldots, Y_k(p)) \\
&\quad + \varpi^* \mu_Q\left(\tilde{X}_1(q), \ldots, \tilde{X}_n(q)\right) \eta(Y_1(p), \ldots, Y_k(p)) \\
&= \mu_Q(X_1(q), \ldots, X_n(q))\, \omega(Y_1(p), \ldots, Y_k(p)) \\
&= \mu_Z.
\end{aligned}$$

Consequently lifts are determined entirely by elements of the quotient space, $\upsilon \in \Upsilon(\varpi : Z \to Q)$. □

Lemma 7. *Let μ_Z be a smooth measure on the total space of a positively-oriented, smooth fiber bundle with μ_Q the corresponding pushforward measure with respect to the projection operator, $\mu_Q = \varpi_* \mu_Z$. If μ_Q is a smooth measure then $\mu_Z = \varpi^* \mu_Q \wedge \xi$ for a unique element of $\xi \in \Xi(\varpi : Z \to Q)$.*

Proof. If the pushforward measure, μ_Q, is smooth then it must satisfy

$$\int_B \mu_Q = \int_{\varpi^{-1}(B)} \mu_Z.$$

Employing a trivializing cover over $\varpi^{-1}(B)$, we can expand the integral over the total space as

$$\begin{aligned}
\int_{\varpi^{-1}(B)} \mu_Z &= \sum_\alpha \int_{\varpi^{-1}(B) \cap (\mathcal{U}_\alpha \times F)} \rho_\alpha \, \mu_Z \\
&= \sum_\alpha \int_{(B \cap \mathcal{U}_\alpha) \times F} \rho_\alpha \, \mu_Z
\end{aligned}$$

By Theorem 6 there is a unique $\upsilon \in \Upsilon(\varpi : Z \to Q)$ such that $\mu_Z = \varpi^* \mu_Q \wedge \upsilon$ and the integral becomes

$$\begin{aligned}
\int_{\varpi^{-1}(B)} \mu_Z &= \sum_\alpha \int_{(B \cap \mathcal{U}_\alpha) \times F} \rho_\alpha \, \mu_Z \\
&= \sum_\alpha \int_{(B \cap \mathcal{U}_\alpha) \times F} \rho_\alpha \, \varpi^* \mu_Q \wedge \upsilon \\
&= \sum_\alpha \int_{(B \cap \mathcal{U}_\alpha)} \rho_\alpha \left[\int_F \iota_q^* \upsilon\right] \mu_Q \\
&= \int_B \rho_\alpha \left[\int_{Z_q} \iota_q^* \upsilon\right] \mu_Q.
\end{aligned}$$

Because μ_Q is σ-finite and $\int_{Z_q} \iota_q^* \upsilon$ is finite, the pushforward condition is satisfied if and only if

$$\int_{Z_q} \iota_q^* \upsilon = 1, \forall q \in Q,$$

which is satisfied if and only if $\upsilon \in \Xi(\varpi : Z \to Q) \subset \Upsilon(\varpi : Z \to Q)$.

Consequently there exists a unique $\xi \in \Xi(\varpi : Z \to Q)$ that lifts the pushforward measure of μ_Z back to μ_Z.

□

Lemma 10. *Let μ_Z be a smooth measure on the total space of a positively-oriented, smooth fiber bundle whose pushforward measure with respect to the projection operator is smooth, with $U \subset Q$ any neighborhood of the base space that supports a local frame. Within $\varpi^{-1}(U)$, the element $\xi \in \Xi(\varpi : Z \to Q)$*

$$\xi = \frac{\left(\tilde{X}_1, \ldots, \tilde{X}_n\right) \lrcorner \mu_Z}{\mu_Q(X_1, \ldots, X_n)},$$

defines the regular conditional probability measure of μ_Z with respect to the projection operator, where $(X_1, \ldots, X_n)$ is any positively-oriented frame in U satisfying

$$\mu_Q(X_1, \ldots, X_n) < \infty, \forall q \in U$$

and $\left(\tilde{X}_1, \ldots, \tilde{X}_n\right)$ is any corresponding horizontal lift.

Proof. Consider any positively-oriented frame on the base space, $(X_1, \ldots, X_n)$, along with any choice of horizontal lift, $\left(\tilde{X}_1, \ldots, \tilde{X}_n\right)$, and an ordered k-tuple of vector fields on the total space, $(Y_1, \ldots, Y_k)$, that restricts to a positively-ordered frame in some neighborhood of the fibers, $V \subset \varpi^{-1}(U)$. Because the fiber bundle is oriented, the horizontal lift and the ordered k-tuple define a positively-ordered frame in V,

$$(W_1, \ldots, W_{n+k}) = \left(\tilde{X}_1, \ldots, \tilde{X}_n, Y_1, \ldots, Y_k\right).$$

If the pushforward measure, μ_Q, is smooth then from Lemma 7 we have $\mu_Z = \varpi^* \mu_Q \wedge \xi$ for some unique $\xi \in \Xi(\varpi : Z \to Q)$. Contracting the frame onto these forms gives

$$\begin{aligned} \mu_Z(W_1, \ldots, W_{n+k}) &= (\varpi^* \mu_Q \wedge \xi)(W_1, \ldots, W_{n+k}) \\ &= \mu_Q(X_1, \ldots, X_n)\, \xi(Y_1, \ldots, Y_k). \end{aligned}$$

Given the positive orientations of the frames and the forms, each term is strictly positive and provided that $\mu_Q(X_1, \ldots, X_n)$ is finite for all $q \in U$ we can divide to give,

$$\xi(Y_1, \ldots, Y_k) = \frac{\mu_Z(W_1, \ldots, W_{n+k})}{\mu_Q(X_1, \ldots, X_n)}.$$

Finally, because ξ is invariant to the addition of horizontal k-forms this implies that within U

$$\xi = \frac{\left(\tilde{X}_1, \ldots, \tilde{X}_n\right) \lrcorner \mu_Z}{\mu_Q(X_1, \ldots, X_n)}.$$

□

Lemma 12. *Let (M, Ω, H) be a Hamiltonian system with the finite and smooth canonical measure, $\mu = e^{-\beta H}$. The microcanonical distribution on the level set $H^{-1}(E)$,*

$$\pi_{H^{-1}(E)} = \frac{v \lrcorner \Omega}{\int_{H^{-1}(E)} \iota_E^* \left(v \lrcorner \Omega \right)},$$

is invariant to the corresponding Hamiltonian flow restricted to the level set, $\phi_t^H \big|_{H^{-1}(E)}$.

Proof. By construction the global flow preserves the canonical distribution,

$$\begin{aligned} \pi &= \left(\phi_t^H \right)_* \pi \\ &= \left(\phi_t^H \right)_* \left(\pi_{H^{-1}(E)} \wedge \pi_E \right) \\ &= \left(\left(\phi_t^H \right)_* \pi_{H^{-1}(E)} \right) \wedge \left(\left(\phi_t^H \right)_* \pi_E \right). \end{aligned}$$

Because the Hamiltonian is itself invariant to the flow we must have

$$\left(\left(\phi_t^H \right)_* \pi_E \right) = \pi_E$$

and

$$\pi = \left(\left(\phi_t^H \right)_* \pi_{H^{-1}(E)} \right) \wedge \pi_E.$$

From Lemma 7, however, the regular conditional probability measure in the decomposition must be unique, hence

$$\begin{aligned} \iota_q^* \pi_{H^{-1}(E)} &= \iota_q^* \left(\left(\phi_t^H \right)_* \pi_{H^{-1}(E)} \right) \\ &= \left(\phi_t^H \circ \iota_q \right)_* \pi_{H^{-1}(E)} \\ &= \left(\iota_q \circ \phi_t^H \big|_{H^{-1}(E)} \right)_* \pi_{H^{-1}(E)} \\ &= \left(\phi_t^H \big|_{H^{-1}(E)} \right)_* \iota_q^* \pi_{H^{-1}(E)}, \end{aligned}$$

as desired.

□

References

Amari, S.-i. and Nagaoka, H. (2007). *Methods of Information Geometry*. American Mathematical Soc.

Baez, J. C. and Muniain, J. P. (1994). *Gauge Fields, Knots, and Gravity*. World Scientific Singapore.

Beskos, A., Pinski, F., Sanz-Serna, J. and Stuart, A. (2011). Hybrid Monte Carlo on Hilbert Spaces. *Stochastic Processes and their Applications* **121** 2201–2230.

Beskos, A., Pillai, N., Roberts, G., Sanz-Serna, J.-M. and Stuart, A. (2013). Optimal Tuning of the Hybrid Monte Carlo Algorithm. *Bernoulli* **19** 1501–1534.

Betancourt, M. (2013a). Generalizing the No-U-Turn Sampler to Riemannian Manifolds. *ArXiv e-prints* **1304.1920**.

Betancourt, M. (2013b). A General Metric for Riemannian Hamiltonian Monte Carlo. In *First International Conference on the Geometric Science of Information* (F. Nielsen and F. Barbaresco, eds.). *Lecture Notes in Computer Science* **8085**. Springer.

Betancourt, M. (2014). Adiabatic Monte Carlo. *ArXiv e-prints* **1405.3489**.

Betancourt, M. and Stein, L. C. (2011). The Geometry of Hamiltonian Monte Carlo. *ArXiv e-prints*.

Burrage, K., Lenane, I. and Lythe, G. (2007). Numerical Methods for Second-Order Stochastic Differential Equations. *SIAM Journal on Scientific Computing* **29** 245–264.

Byrne, S. and Girolami, M. (2013). Geodesic Monte Carlo on Embedded Manifolds. *Scandinavian Journal of Statistics* **40**.

Caflisch, R. E. (1998). Monte Carlo and Quasi-Monte Carlo Methods. *Acta numerica* **7** 1–49.

Cannas da Silva, A. (2001). *Lectures on Symplectic Geometry*. Springer.

Censor, A. and Grandini, D. (2014). Borel and Continuous Systems of Measures. *To appear in Rocky Mountain Journal of Mathematics*.

Chang, J. T. and Pollard, D. (1997). Conditioning as Disintegration. *Statistica Neerlandica* **51** 287–317.

Cotter, S., Roberts, G., Stuart, A., White, D. et al. (2013). MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster. *Statistical Science* **28** 424–446.

Cowling, B., Freeman, G., Wong, J., Wu, P., Liao, Q., Lau, E., Wu, J., Fielding, R. and Leung, G. (2012). Preliminary Inferences on the Age-Specific Seriousness of Human Disease Caused by Avian Influenza A (H7N9) Infections in China, March to April 2013. *European Communicable Disease Bulletin* **18**.

Dawid, A. P., Stone, M. and Zidek, J. V. (1973). Marginalization Paradoxes in Bayesian and Structural Inference. *Journal of the Royal Statistical Society. Series B. Methodological* **35** 189–233.

Diaconis, P. and Freedman, D. (1999). Iterated Random Functions. *SIAM review* **41** 45–76.

Draganescu, A., Lehoucq, R. and Tupper, P. (2009). Hamiltonian Molecular Dynamics for Computational Mechanicians and Numerical Analysts Technical Report No. 2008-6512, Sandia National Laboratories.

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

Fang, Y., Sanz-Serna, J.-M. and Skeel, R. D. (2014). Compressible Generalized Hybrid Monte Carlo. *The Journal of chemical physics* **140** 174108.

Federer, H. (1969). *Geometric Measure Theory*. Springer New York.

Fixman, M. (1978). Simulation of Polymer Dynamics. I. General Theory. *The Journal of Chemical Physics* **69** 1527–1537.

Folland, G. B. (1999). *Real Analysis: Modern Techniques and Their Applications*. John Wiley and Sons, Inc, New York.

Frenkel, D. and Smit, B. (2001). *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American statistical association* **85** 398–409.

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **6** 721–741.

Ghitza, Y. and Gelman, A. (2014). The Great Society, Reagan's Revolution, and Generations

of Presidential Voting. *To be submitted.*

Girolami, M. and Calderhead, B. (2011). Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73** 123–214.

Haario, H., Saksman, E. and Tamminen, J. (2001). An Adaptive Metropolis Algorithm. *Bernoulli* 223–242.

Haile, J. M. (1992). *Molecular Dynamics Simulation: Elementary Methods*. John Wiley & Sons, Inc.

Hairer, E., Lubich, C. and Wanner, G. (2006). *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*. Springer, New York.

Halmos, P. R. (1950). *Measure Theory*. D. Van Nostrand Company, Inc., New York, N. Y.

Hoffman, M. D. and Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* **15** 1593–1623.

Holmes, S., Rubinstein-Salzedo, S. and Seiler, C. (2014). Curvature and Concentration of Hamiltonian Monte Carlo in High Dimensions.

Horowitz, A. M. (1991). A Generalized Guided Monte Carlo Algorithm. *Physics Letters B* **268** 247–252.

Husain, S., Vasishth, S. and Srinivasan, N. (2014). Strong Expectations Cancel Locality Effects: Evidence from Hindi. *PLoS ONE* **9** e100986.

Izaguirre, J. A. and Hampton, S. S. (2004). Shadow Hybrid Monte Carlo: An Efficient Propagator In Phase Space Of Macromolecules. *Journal of Computational Physics* **200** 581–604.

Jasche, J., Kitaura, F. S., Li, C. and Ensslin, T. A. (2010). Bayesian Non-Linear Large-Scale Structure Inference of the Sloan Digital Sky Survey Data Release 7. *Monthly Notices of the Royal Astronomical Society* **409** 355–370.

José, J. V. and Saletan, E. J. (1998). *Classical Dynamics: A Contemporary Approach*. Cambridge University Press, New York.

Joulin, A. and Ollivier, Y. (2010). Curvature, Concentration and Error Estimates for Markov Chain Monte Carlo. *The Annals of Probability* **38** 2418–2442.

Kardar, M. (2007). *Statistical Physics of Particles*. Cambridge University Press.

Lan, S., Stathopoulos, V., Shahbaba, B. and Girolami, M. (2012). Lagrangian Dynamical Monte Carlo.

Leão Jr, D., Fragoso, M. and Ruffino, P. (2004). Regular Conditional Probability, Disintegration of Probability and Radon Spaces. *Proyecciones (Antofagasta)* **23** 15–29.

Lee, J. M. (2011). *Introduction to Topological Manifolds*. Springer.

Lee, J. M. (2013). *Introduction to Smooth Manifolds*. Springer.

Leimkuhler, B. and Reich, S. (2004). *Simulating Hamiltonian Dynamics*. Cambridge University Press, New York.

Livingstone, S. and Girolami, M. (2014). Information-Geometric Markov Chain Monte Carlo Methods Using Diffusions. *Entropy* **16** 3074–3102.

Marx, D. and Hutter, J. (2009). *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press.

Meyn, S. P. and Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability*. Cambridge University Press.

Murray, I. and Elliott, L. T. (2012). Driving Markov Chain Monte Carlo with a Dependent Random Stream.

Neal, R. M. (2011). MCMC Using Hamiltonian Dynamics. In *Handbook of Markov Chain Monte*

Carlo (S. Brooks, A. Gelman, G. L. Jones and X.-L. Meng, eds.) CRC Press, New York.

Neal, R. M. (2012). How To View an MCMC Simulation as a Permutation, with Applications to Parallel Simulation and Improved Importance Sampling.

Øksendal, B. (2003). *Stochastic Differential Equations*. Springer.

Ollivier, Y. (2009). Ricci Curvature of Markov Chains on Metric Spaces. *Journal of Functional Analysis* **256** 810–864.

Petersen, K. (1989). *Ergodic Theory*. Cambridge University Press.

Polettini, M. (2013). Generally Covariant State-Dependent Diffusion. *Journal of Statistical Mechanics: Theory and Experiment* **2013** P07005.

Porter, E. K. and Carré, J. (2014). A Hamiltonian Monte Carlo Method for Bayesian Inference of Supermassive Black Hole Binaries. *Classical and Quantum Gravity*.

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer New York.

Roberts, G. O., Rosenthal, J. S. et al. (2004). General State Space Markov Chains and MCMC Algorithms. *Probability Surveys* **1** 20–71.

Roberts, G. O., Gelman, A., Gilks, W. R. et al. (1997). Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms. *The annals of applied probability* **7** 110–120.

Sanders, N., Betancourt, M. and Soderberg, A. (2014). Unsupervised Transient Light Curve Analysis Via Hierarchical Bayesian Inference. *ArXiv e-prints*.

Schofield, M. R., Barker, R. J., Gelman, A., Cook, E. R. and Briffa, K. (2014). Climate Reconstruction Using Tree-Ring Data. *Journal of the American Statistical Association (under revision)*.

Schutz, B. (1980). *Geometrical Methods of Mathematical Physics*. Cambridge University Press, New York.

Simmons, D. (2012). Conditional Measures and Conditional Expectation; Rohlin's Disintegration Theorem. *Discrete and Dynamical Systems* **32** 2565–2582.

Sohl-Dickstein, J., Mudigonda, M. and DeWeese, M. (2014). Hamiltonian Monte Carlo Without Detailed Balance. In *Proceedings of the 31st International Conference on Machine Learning* 719–726.

Souriau, J.-M. (1997). *Structure of Dynamical Systems: A Symplectic View of Physics*. Springer.

Sutherland, D. J., Póczos, B. and Schneider, J. (2013). Active Learning and Search on Low-Rank Matrices. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '13* 212–220. ACM.

Tang, Y., Srivastava, N. and Salakhutdinov, R. (2013). Learning Generative Models with Visual Attention. *ArXiv e-prints*.

Stan Development Team (2014). Stan: A C++ Library for Probability and Sampling, Version 2.5.

Terada, R., Inoue, S. and Nishihara, G. N. (2013). The Effect of Light and Temperature on the Growth and Photosynthesis of Gracilariopsis Chorda (Gracilariales, Rhodophtya) from Geographically Separated Locations of Japanowth and Photosynthesis of Gracilariopsis Chorda (Gracilariales, Rhodophtya) from Geographically Separated Locations of Japan. *Journal of Applied Phycology* 1-10.

Tierney, L. (1998). A Note on Metropolis-Hastings Kernels for General State Spaces. *The Annals of Applied Probability* **8** 1–9.

Tjur, T. (1980). *Probability Based on Radon Measures*. Wiley Chichester.

Wang, Z., Mohamed, S. and de Freitas, N. (2013). Adaptive Hamiltonian and Riemann Manifold Monte Carlo. In *Proceedings of the 30th International Conference on Machine Learning*

(ICML-13) 1462–1470.

Wang, H., Mo, H. J., Yang, X., Jing, Y. P. and Lin, W. P. (2014). ELUCID - Exploring the Local Universe with Reconstructed Initial Density field I: Hamiltonian Markov Chain Monte Carlo Method with Particle Mesh Dynamics‘. *ArXiv e-prints*.

Weber, S., Carpenter, B., Lee, D., Bois, F. Y., Gelman, A. and Racine, A. (2014). Bayesian Drug Disease Model with Stan - Using Published Longitudinal Data Summaries in Population Models. *Abstracts of the Annual Meeting of the Population Approach Group in Europe* **23**.

Weinstein, A. (1983). The Local Structure of Poisson Manifolds. *Journal of Differential Geometry* **18** 523–557.

Zaslavsky, G. M. (2005). *Hamiltonian Chaos and Fractional Dynamics*. Oxford University Press Oxford.