HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Master’s thesis

Mathematics

The differential geometry of Markov transitions in Hamiltonian Monte Carlo

Jussi Penttinen

Supervisor: Antti Honkela

Matematiikan ja tilastotieteen maisteriohjelma / Magisterprogram i matematik och statistik / Master’s Programme in Mathematics and Statistics

Matemaattis-luonnontieteellinen tiedekunta / Matematisk-naturvetenskapliga fakulteten / Faculty of Science

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Laitos — Institution — Department
Mathematics and statistics	Faculty of science

Tekijä — Författare — Author

Jussi Penttinen

Työn nimi — Arbetets titel — Title

The differential geometry of Markov transitions in Hamiltonian Monte Carlo

Oppiaine — Läroämne — Subject

Mathematics

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Master's thesis	February 2021	43

Tiivistelmä — Referat — Abstract

HMC is a computational method build to efficiently sample from a high dimensional distribution. Sampling from a distribution is typically a statistical problem and hence a lot of works concerning Hamiltonian Monte Carlo are written in the mathematical language of probability theory, which perhaps is not ideally suited for HMC, since HMC is at its core differential geometry.

The purpose of this text is to present the differential geometric tool's needed in HMC and then methodically build the algorithm itself. Since there is a great introductory book to smooth manifolds by Lee and not wanting to completely copy Lee's work from his book, some basic knowledge of differential geometry is left for the reader.

Similarly, the author being more comfortable with notions of differential geometry, and to cut down the length of this text, most theorems connected to measure and probability theory are omitted from this work.

The first chapter is an introductory chapter that goes through the bare minimum of measure theory needed to motivate Hamiltonian Monte Carlo. Bulk of this text is in the second and third chapter. The second chapter presents the concepts of differential geometry needed to understand the abstract build of Hamiltonian Monte Carlo. Those familiar with differential geometry can possibly skip the second chapter, even though it might be worth while to at least flip through it to fill in on the notations used in this text.

The third chapter is the core of this text. There the algorithm is methodically built using the groundwork laid in previous chapters. The most important part and the theoretical heart of the algorithm is presented here in the sections discussing the lift of the target measure.

The fourth chapter provides brief practical insight to implementing HMC and also discusses quickly how HMC is currently being improved.

Avainsanat — Nyckelord — Keywords

Hamiltonian Monte Carlo, differential geometry, Monte Carlo, Markov chains.

Säilytyspaikka — Förvaringsställe — Where deposited

Helda achive

Muita tietoja — Övriga uppgifter — Additional information

Contents

Chapter 1

Introduction

Hamiltonian Monte Carlo is a computational method that was born in practical applications, and then further development with empirical experiments spearheading the research [4]. The method is now some thirty years old and theoretical results backing up the empirical successes are finally catching up [4].

The slow theoretical progress of HMC arise not only from complexity of applied probability theory necessary for general state space Markov chains, but also from high levels of abstractness intrinsic to differential geometry.

This text focuses on the differential geometric tools needed in Hamiltonian Monte Carlo and shows how those tools are used to build the Markov transitions. The measure theory of Markov chains is not covered in this text and hence unfortunately, this is not a complete guide to Hamiltonian Monte Carlo.

1.1 Prerequisites

To keep this text within readable length, some familiarity with differential geometry [1] and basic knowledge of Markov chains [3] is assumed.

For those not fully comfortable with differential geometry I heartily recommend the introductory book by Lee [1]. Even for those more experienced in differential geometry, it serves as a good backup, because most of the results in this text are more thoroughly covered by Lee in his book.

The most important notions of differential geometry that are not covered in this text are tangent vectors, covectors, differential forms, tensor product, wedge product, vector fields and flows.

Since applied probability theory in itself is very complex matter [7], care-fully going through both the measure theory of Markov chains and differential geometry of Hamiltonian Monte Carlo would probably require a whole book. For that reason, the results and theorems connected to probability theory are mostly omitted from this text.

In [2], [4] the authors go through the effort of building some bridges between differential geometry and measure theory and for that reason I would recommend [2], [4] for those more familiar with measure theory.

For those who are interested in other branches of mathematics closely related to Hamiltonian Monte Carlo, I would suggest studying dynamical systems and ergodic theory [5].

1.2 Motivation for Markov chain Monte Carlo methods

Problems in statistics usually boil down to computing expectations [4] $E_\mu[f]$ of some function f on the sample space M with respect to a distribution μ.

If the sample space is for example $\mathbb{R}^n$, the distribution μ can be identified as smooth probability density function and the expectations are calculated by a simple looking integration

$$E_\mu[f] = \int_{\mathbb{R}^n} f(x)\mu(x)\, \mathrm{d}x^1 ... \mathrm{d}x^n. \tag{1.1}$$

But even in this simpler case of $\mathbb{R}^n$ the integral may not be analytically solvable and numerical methods become the only option. Unfortunately however, despite the ever increasing computational power of computers, the numerical methods are still bounded by our finite time and patience and hence, calculating such integrals quickly becomes impossible.

Now imagine that there is a sufficiently large but finite collection of points $\{p_i\}_{i \in I_k}$, $I_k = \{1, 2, ..., k\}$ that somehow corresponds to the distribution μ in the sense that, averaging function f over the points eventually creates exactly the expectation

$$E_\mu[f] = \lim_{k \to \infty} \sum_{i \in I_k} \frac{1}{k} f(p_i). \tag{1.2}$$

Computing (1.2) to some sufficiently large k is significantly easier than numerically integrating (1.1). Therefore there are high interests in methods that create accurate enough samples of μ in reasonable time and Markov chain Monte Carlo methods, includin Hamiltonian Monte Carlo, are indeed the tools designed answers this problem.

1.3 Markov kernels and transitions

To properly set up for Hamiltonian Monte Carlo, some knowledge of Markov kernels in continuous state spaces is needed.

Assume that $(X, \mathcal{A}, \mu)$ is a measurable space with a sigma-algebra $\mathcal{A}$ and a given probability measure μ. A *Markov kernel* τ is a map

$$\tau \colon X \times \mathcal{A} \to [0, 1]$$

that has the following two properties [2]:

1. τ is measurable function in the first argument

$$\tau(\cdot, A) \colon X \to [0, 1], \qquad \forall A \in \mathcal{A}.$$

2. τ is probability measure in the second argument

$$\tau(q, \cdot) \colon \mathcal{A} \to [0, 1], \qquad \forall q \in X.$$

One intuition into Markov kernels is that $\tau(q, A)$ is the conditional probability of transitioning from point q to a point $q' \in A$.

Easy corollary of the second property is that a Markov kernel is also a map from X to the space of probability measures $\mathcal{P}(X)$ on X i.e.

$$\tau(q, \cdot) \in \mathcal{P}(X), \qquad \forall q \in X,$$

which further implies that Markov kernels give instructions how to sample a new point q' if starting from point q.

Not all Markov Kernels are created equal however, because taking the given measure μ, any $A \in \mathcal{A}$ and integrating the Markov kernel over the entire X

$$\mu'(A) = \int_X \tau(q, A)\mu(\mathrm{d}q), \qquad A \in \mathcal{A} \tag{1.3}$$

gives some measure $\mu'(A)$ to the set A. If it holds that $\mu'(A) = \mu(A)$ for every $A \in \mathcal{A}$, the Markov kernel is said to preserve the measure μ.

Formula (1.3) defines a map called *Markov transition* from the space of probability measures $\mathcal{P}(X)$ to itself [2]:

$$\begin{aligned} \mathcal{T} \colon \mathcal{P}(X) &\to \mathcal{P}(X) \\ \mathcal{T}\mu\,(A) &= \int_X \tau(q, A)\mu(\mathrm{d}q). \end{aligned} \tag{1.4}$$

If the Markov kernel is aperiodic, irreducible, Harris recurrent, and preserves the target measure as in (1.3), multiple Markov transitions will create a Markov chain that will eventually explore whole of μ and fulfil Equation (1.2) [2], [3].

1.4 Markov kernel induced from measure preserving maps

To make the awkward transition from probability theory to differential geometry sufficient for this text, consider measure preserving maps [5].

Generally, if $(X, \mathcal{A}, \mu)$ is a measurable space, a map $t\colon X \to X$ is called measure preserving if $\mu(A) = \mu(t^{-1}(A))$ for every $A \in \mathcal{A}$.

Assume that T is a space of continuous bijective measure preserving-maps on measurable space $(X, \mathcal{A}, \mu)$ i.e.

$$T = \{t\colon X \to X \,\big|\, \mu(A) = \mu(t^{-1}(A)),\ \forall A \in \mathcal{A}\}.$$

Given some σ-algebra $\mathcal{G}$ over T and some probability measure g on $\mathcal{G}$, $(T, \mathcal{G}, g)$ defines a probability space [2], which induces a Markov kernel [6] as an iterated random function [7]:

$$\tau(q, A) \equiv \int_T \mathbb{I}_A(t(q))\, g(\mathrm{d}t) \tag{1.5}$$

$$\mathbb{I}_A(t(q)) \begin{cases} = 1 \text{ if } t(q) \in A \\ = 0 \text{ if } t(q) \notin A. \end{cases}$$

In this context, the kernel (1.5) containing an integral over a space of functions is still very abstract, especially because the measure g is left completely unspecified. But as the definition for Hamiltonian Monte Carlo solidifies in Section 3.5, the measure g gets defined quite naturally and the integral turns into a more comprehensive integral over a vector space.

1.5 Intuition behind Hamiltonian Monte Carlo

Assume M is a smooth n-manifold, the given measure μ_M in terms of differential geometry is then a non-vanishing differential n-form with the property that

$$\int_M \mu_M = 1.$$

Since Euclidean spaces are manifolds, to help with intuition, it is often easier to think of M as $\mathbb{R}^n$, the measure μ_M as a positive smooth function $\mu_M : \mathbb{R}^n \to \mathbb{R}$ with property $\mu_M(\mathbb{R}^n) = 1$, and the measure $\mu_M(A)$ defined as

$$\mu_M(A) = \int_A \mu_M.$$

The goal is to build a family of measure-preserving functions so that from each point $q \in M$, there exists preferably multiple functions taking different values in M.

In Hamiltonian Monte Carlo, this is done by building a Hamiltonian flow generated by a Hamiltonian vector field, which is a particular, well-known vector field introduced later. The Hamiltonian flow is by construction measure-preserving, but the problem is that, the Hamiltonian vector fields and flows require so called sympletic structure which is defined only on manifolds with even number of dimension.

The simplest way of ensuring even dimensions is by doubling the n-dimensions of the original space. Luckily in differential geometry there is two very fundamental ways of creating manifolds with double of the dimensions of the original manifold. Namely, the tangent bundle and the cotangent bundle, from which the latter proves to be extremely useful in HMC, because of the natural connection between covectors and differential forms.

For example, in the simpler case of $\mathbb{R}^n$ and $\mathbb{R}^n \times \mathbb{R}^n$, it happens that after defining the Hamiltonian flow on $\mathbb{R}^n \times \mathbb{R}^n$, every point on fibre $q \times \mathbb{R}^n$ defines a measure preserving function on $\mathbb{R}^n \times \mathbb{R}^n$, which then gets projected back on to the original $\mathbb{R}^n$ and thus creates the family of measure-preserving functions from $\mathbb{R}^n$ to $\mathbb{R}^n$ in question.

Creation of the Hamiltonian vector field takes a lot of theoretical groundwork, in which, the most crucial part is the lifting of the information of the measure μ_M to the $2n$-manifold so that the Hamiltonian flow that preserves μ_M can be created.

Chapter 2

The building blocks

This section is an overview of the differential geometry needed in Hamiltonian Monte Carlo.

Assume that the sample space M is a connected n-dimensional smooth manifold. Since μ_M is a probability measure, it must be non-vanishing and

$$\int_M \mu_M = 1.$$

Therefore μ_M is an orientation form on M and M is an orientable manifold.

Note that the *Einstein summation convention*

$$a_i x^i = \sum_{i=1}^{n} a_i x^i,$$

when the index i is left unspecified, is used throughout this text as it is also widely used in differential geometry in general.

2.1 Vector bundle

A vector bundle of rank k over M is a joint structure of two topological spaces E and M, and a continuous surjective map $\pi : E \to M$ between them, satisfying two conditions:

1. The set $\pi^{-1}(q) = E_q \subset E$ has the structure of a k-dimensional vector space. The set E_q is called the *fibre* of E over point q.

2. For every point $q \in M$ there exists a neighbourhood $U \subset M$ and a homeomorphism $\Phi : \pi^{-1}(U) \to U \times \mathbb{R}^k$, such that diagram in Figure (2) commutes and the restriction of Φ on any fibre E_q is a linear isomorphism from E_q to $q \times \mathbb{R}^k \cong \mathbb{R}^k$.

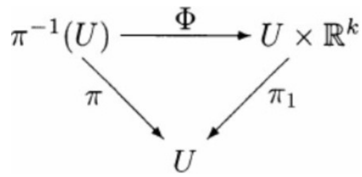

Figure 2.1: π_1 is projection on the first factor. Figure is from [1]

The homeomorphism $\Phi : \pi^{-1}(U) \to U \times \mathbb{R}^k$ is called a *local trivialisation* of E over U. The space E is called the *total space* of the vector bundle, M is referred as its *base space* and the map π is called its *projection.*

The mental image of a vector bundle is that locally it looks like as if to every point of the base is attached a nice k-dimensional euclidean vector space but from global view that attachment can be somehow twisted.

It is possible that a single trivialisation encompasses the whole base space i.e. there is a homeomorphism $\Phi : \pi^{-1}(M) \to M \times \mathbb{R}^k$ and then the total space itself is homeomorphic to $M \times \mathbb{R}^k$. In this case the vector bundle is said to be *trivial bundle* and the local trivialisation a *global trivialisation.*

If both, the total space and the base space are smooth manifolds the local trivialisations can be chosen to be diffeomorphisms instead of homeomorphisms and then the vector bundle is called a *smooth vector bundle.*

2.2 Smooth cotangent bundle as a smooth vector bundle and as a smooth manifold

Let $(\mathcal{U}, x)$ be a smooth chart on a smooth manifold M and $(\frac{\partial}{\partial x^1}, ..., \frac{\partial}{\partial x^n})$ the coordinate vector frame on the aforementioned open set $\mathcal{U}$.

The corresponding covector frame $(\mathrm{d}x^1, ..., \mathrm{d}x^n)$ is defined as duals

$$\mathrm{d}x^1 \colon T_qM \to \mathbb{R}, \quad \forall q \in \mathcal{U}$$

so that

$$\mathrm{d}x^i \left(\frac{\partial}{\partial x^j} \right) = \frac{\partial}{\partial x^j}(x^i) = \delta_i^j. \tag{2.1}$$

The covector frame brings the same vector space structure of the tangent space T_qM to the corresponding cotangent space T_q^*M and the cotangent bundle is then defined as the disjoint union of the cotangent spaces:

$$T^*M = \bigsqcup_{q \in M} T_q^*M.$$

This also makes T^*M a smooth manifold in itself, since every point in the cotangent bundle can be identified as a pair (q, φ), where q is a point in M and φ a covector in T_q^*M.

Therefore one can always build smooth charts χ_i near any point of T^*M as

$$\begin{aligned} \chi_i(q, \varphi) &= (q^1, ..., q^n, p_1, ..., p_n) \\ &= (q^i, p_i), \end{aligned} \tag{2.2}$$

where (q^i) is the coordinate rerepresentation of q and $p_i \mathrm{d}q^i$ is the coordinate rerepresentation of φ. These coordinates (2.2) are called the *standard coordinates* on T^*M.

This does not completely prove that the cotangent bundle is a smooth manifold, as there are still other conditions left to check. The formal proof is thoroughly done in [1] and since it is somewhat unsurprising after the introduction of the standard coordinates, the rest of the proof is omitted from this text.

As the name suggests the cotangent bundle also has a natural vector bundle structure where T^*M identified as the total space, M as the base space and the projection $\pi : T^*M \to M$ is simply

$$\pi(q, \varphi) = q.$$

A fibre $(T^*M)_q$ over q has the vector space structure of the cotangent space T_q^*M.

2.3 Symplectic manifold

In essence a symplectic manifold is an even-dimensional smooth manifold that has a smooth non-degenerate closed 2-form defined on it. A tensor is called *non-degenerate* if

$$\omega(v_1, v_2) = 0$$

for all vectors v_2, implies that

$$v_1 = 0.$$

The symplectic form is sometimes also referred as the symplectic structure of the manifold.

The even number of dimension is a clue that, of course, the cotangent bundle is also a symplectic manifold. The natural or canonical symplectic structure of the cotangent bundle is still unspecified however, and that is the subject of the next section.

2.4 Cotangent bundle as a symplectic manifold

The projection π along with the fact that every point in the cotangent bundle can be identified as a covector φ in the corresponding cotangent space T_q^*M, enables the definition of a canonical 1-form τ in T^*M called the *tautological 1-form*.

For every point $(q,\varphi) \in T^*M$, $\tau_{(q,\varphi)}$ is defined as the pull-back of the projection π

$$\tau_{(q,\varphi)} = \pi^*\varphi$$

i.e. for a tangent vector $X \in T_{(q,\varphi)}(T^*M)$,

$$\tau_{(q,\varphi)}X = \pi^*\varphi(X) = \varphi(\pi_* X).$$

In itself, the tautological form defined like this is not too remarkable. However, as the next theorem shows, the tautological form is a smooth 1-form and most notably, $\omega = -\mathrm{d}\tau$ defines a symplectic form on T^*M.

Before proving that ω is a sympletic form, little bit knowledge of *exterior derivative* d is needed. The following definition from [1] is sufficient for the purpose's of this text.

On any smooth manifold M there is a unique linear map $\mathrm{d} : \Omega^k(M) \to \Omega^{k+1}(M)$ for each $k \geq 0$ satisfying the following conditions:

1. If f is a smooth function (a 0-form), then $\mathrm{d}f$ is the differential of f, defined by
$$\mathrm{d}f(X) = X(f)$$
for any vector field X on M.

2. If $\alpha \in \Omega^k(M)$ and $\beta \in \Omega^l(M)$, then
$$\mathrm{d}(\alpha \wedge \beta) = \mathrm{d}\alpha \wedge \beta + (-1)^k \alpha \wedge \mathrm{d}\beta.$$

3. For any $\alpha \in \Omega^k(M)$
$$\mathrm{d}(\mathrm{d}\alpha) = 0. \tag{2.3}$$

Theorem 2.4.1. *Let M be a smooth manifold. The tautological 1-form is smooth, and $\omega = -d\tau$ is a symplectic form on the covector bundle T^*M [1].*

The following proof is almost identical to the one found in [1].

Proof. Let (q^i) be any smooth coordinates on M and let (q^i, p_i) denote the corresponding standard coordinates on T^*M. Then the coordinate rerepresentation of τ is

$$\begin{aligned}\tau_{(q,p)} &= \pi_q^*(p_i\,\mathrm{d}q^i) \quad \in \Omega^1(T^*M)\\ &= p_i\,\mathrm{d}q^i \quad \in \Omega^1(M),\end{aligned}$$

which is smooth, and then τ is also smooth.

By definition $\omega = \mathrm{d}\tau$, ω is exact and by (2.3) it holds that

$$\mathrm{d}\omega = \mathrm{d}(-\mathrm{d}\tau) = 0,$$

and therefore ω is closed.

The non-degeneracy condition can be checked by first opening the definition of ω by a straightforward calculation

$$\begin{aligned}\omega &= -\mathrm{d}\tau = -\mathrm{d}(p_i\,\mathrm{d}q^i) = -\sum_{i=1}^{n}\mathrm{d}p_i \wedge \mathrm{d}q^i\\ \omega &= \sum_{i=1}^{n}\mathrm{d}q^i \wedge \mathrm{d}p_i.\end{aligned} \tag{2.4}$$

Now the action of ω on any basis vectors of T^*M, which are the form $(\partial q_1, ..., \partial q_n, \partial p_1, ..., \partial p_n)$ is

$$\begin{aligned}\omega(\partial q_i, \partial p_j) &= -\omega(\partial p_j, \partial q_i) = \delta_{ij}\\ \omega(\partial q_i, \partial q_j) &= \omega(\partial p_i, \partial p_j) = 0.\end{aligned}$$

Suppose that $V = \alpha^i\,\partial q_i + \beta^i\,\partial p_i \in T^*M$ and $\omega(V, W) = 0$ for all $W \in T^*M$. Then it would hold that

$$\begin{aligned}\omega(V, \partial q_i) &= -\beta^i = 0\\ \omega(V, \partial p_i) &= \alpha^i = 0,\end{aligned}$$

which implies that $V = 0$ and ω is non-degenerate and hence a symplectic form on T^*M. □

The symplectic form (2.4) is called the *canonical symplectic form* on T^*M.

The canonical symplectic form (2.4) along with the standard coordinates (q^i, p_i) fulfil the Darboux theorem [2], meaning that near every point of the cotangent bundle there are smooth standard coordinates in which the canonical symplectic form has the coordinate representation (2.4).

Using the convenient standard coordinates a *canonical volume form* can be defined as

$$\Omega = \mathrm{d}q_1 \wedge ... \wedge \mathrm{d}q_n \wedge \mathrm{d}p_1 \wedge ... \wedge \mathrm{d}p_n. \tag{2.5}$$

2.5 Hamiltonian system

The symplectic form is used to define one last piece of important machinery called the *Hamiltonian vector field*.

One can think of the symplectic form ω also as a map from the tangent bundle to the cotangent bundle i.e.

$$\begin{aligned} &\tilde{\omega}\colon TM \to T^*M \\ &\tilde{\omega}(v) = v \lrcorner\, \omega = \omega(v, \cdot), \quad v \in TM. \end{aligned}$$

Since the symplectic form ω is smooth and non-degenerate by definition, the map $\tilde{\omega}_q$ is continuous injection in every fibre T_qM. Furthermore, because $\dim T_qM = \dim T_q^*M$ and $\tilde{\omega}_q$ is linear map with a null-space of $\{0\}$, $\tilde{\omega}_q$ is also a surjection.

Therefore there exists an inverse $\tilde{\omega}^{-1} : T^*M \to TM$ such that for any $H \in C^\infty(T^*M)$ a *Hamiltonian vector field* can be defined:

$$X_H = \tilde{\omega}^{-1}(\mathrm{d}H),$$

where the defining function H is called the *Hamiltonian function*.

In other words, X_H is a vector field for which the equation

$$X_H \lrcorner\, \omega = \mathrm{d}H$$

holds, and for any vector field V on the total space T^*M

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

where $V(H)$ can be thought of as a directional derivative of function H in the direction of vector V.

Smooth manifold along with a symplectic form and a Hamiltonian function is called a *Hamiltonian system* and it has few useful properties.

Firstly, since X_H is smooth vector field by definition, then by the fundamental theorem on flows [1], there exists a maximal *Hamiltonian flow* $\theta^H_{(t,p)}$ generated by the Hamiltonian vector field.

Moreover, if one checks how the Hamiltonian function behaves on the Hamiltonian flow, one would find that

$$\mathcal{L}_{X_H} H = X_H H = \mathrm{d}H X_H = \omega(X_H, X_H) = 0, \tag{2.6}$$

which implies that the Hamiltonian function is constant on the Hamiltonian flow and the Hamiltonian vector field is tangent to the level sets consisting of the regular points of the Hamiltonian function.

Secondly, assume that X_H is non-vanishing. Then by Global Frobenius Theorem [1] the maximal integral curves *foliate* the manifold T^*M into disjoint 1-dimensional immersed submanifolds so that

$$T^*M = \bigsqcup_{i \in I} \gamma_i(t), \tag{2.7}$$

where γ_i is a maximal integral curve.

The foliation (2.7) of the cotangent bundle is a confirmation that the cotangent bundle, and especially the base manifold, can be completely explored by using the integral curves of the Hamiltonian vector field. But as a side note, the complete foliation of T^*M is not needed to be able to explore M completely, as the whole of T^*M is not needed for the projection π to be surjective.

The points where a vector field vanishes are called the critical points of the vector field and the critical points of the Hamiltonian vector field are tied to the Hamiltonian function in the following way.

$$\begin{aligned} X_{H(p)} = 0 &\Leftrightarrow \mathrm{d}H(p) = 0 \\ &\Leftrightarrow VH(p) = 0 \quad \forall V \in T_p(T^*M). \end{aligned}$$

The flow and critical points of a vector field are extensively studied in differential equations. The specific requirements of Hamiltonian Monte Carlo are discussed later in this text.

2.5.1 Local expression for Hamiltonian vector field

The global definition of the Hamiltonian vector field is somewhat abstract and moreover, to be able to use this algorithm in any practical setting, a local expression is needed. Luckily, the standard coordinates provide an excellent medium for local calculations.

First of all, every smooth vector field has the standard coordinate representation

$$X_H = \sum_{i=1}^{n} A^i \frac{\partial}{\partial q^i} + B^i \frac{\partial}{\partial p^i} \tag{2.8}$$

for some smooth functions A^i and B^i.

Then by opening the wedge product

$$\begin{aligned}
\omega &= \mathrm{d}q_j \wedge \mathrm{d}p_j \\
&= \sum_{j=1}^{n} \mathrm{d}q_j \wedge \mathrm{d}p_j \\
&= \sum_{j=1}^{n} \frac{1}{2!} \operatorname{Alt}(\mathrm{d}q_j \otimes \mathrm{d}p_j) \\
&= \sum_{j=1}^{n} \frac{1}{2!} \frac{(1+1)!}{1!1!} (\mathrm{d}q_j \otimes \mathrm{d}p_j - \mathrm{d}p_j \otimes \mathrm{d}q_j) \\
&= \sum_{j=1}^{n} (\mathrm{d}q_j \otimes \mathrm{d}p_j - \mathrm{d}p_j \otimes \mathrm{d}q_j),
\end{aligned}$$

the actual interior multiplication $X_H \lrcorner\, \omega$ with (2.8) can be calculated

$$\begin{aligned}
X_H \lrcorner\, \omega = X_H \lrcorner \sum_{j=1}^{n} (\mathrm{d}q_j \otimes \mathrm{d}p_j - \mathrm{d}p_j \otimes \mathrm{d}q_j) \\
&= \sum_{j=1}^{n} (\mathrm{d}q_j(X_H)\mathrm{d}p_j - \mathrm{d}p_j(X_H)\mathrm{d}q_j) \\
&= \sum_{j=1}^{n} \Bigg(\mathrm{d}q_j \Bigg(\sum_{i=1}^{n} A^i \frac{\partial}{\partial q^i} + B^i \frac{\partial}{\partial p^i} \Bigg) \mathrm{d}p_j \\
&\qquad - \mathrm{d}p_j \Bigg(\sum_{i=1}^{n} A^i \frac{\partial}{\partial q^i} + B^i \frac{\partial}{\partial p^i} \Bigg) \mathrm{d}q_j \Bigg) \\
&= \sum_{j=1}^{n} A^j \mathrm{d}p_j - B^j \mathrm{d}q_j.
\end{aligned}$$

On the other hand, though any covector field has the representation in standard coordinates:

$$dH = \sum_{j=1}^{n} \frac{\partial H}{\partial q^j} dq_j + \frac{\partial H}{\partial p^j} dp_j.$$

So it can be seen that $A_j = \frac{\partial H}{\partial p^j}$ and $B_j = -\frac{\partial H}{\partial q^j}$.

Finally putting it all together, the standard coordinate representation of the Hamiltonian vector field is:

$$X_H = \sum_{j=1}^{n} \frac{\partial H}{\partial p^j}\frac{\partial}{\partial q^j} - \frac{\partial H}{\partial q^j}\frac{\partial}{\partial p^j}.$$

Interestingly enough, if one would try to solve the integral curves of the Hamiltonian vector field, one would notice that that the curves $\gamma(t) = (q^i(t), p^i(t))$ would have to fulfil the *Hamiltonian equations*

$$\begin{aligned}(q^i)'(t) &= \frac{\partial H}{\partial p^i}(q(t), p(t)) \\ (p^i)'(t) &= -\frac{\partial H}{\partial q^i}(q(t), p(t))\end{aligned} \tag{2.9}$$

and hence, one more motivation for the name Hamiltonian vector field.

2.6 Riemannian metric tensor

Even though Riemannian metric is not strictly necessary for the theoretical function of the algorithm, it is still a very useful tool for defining general functions from the cotangent space to real numbers.

Every smooth manifold allows a Riemannian metric [1] and a Riemannian metric tensor field g is a smooth symmetric 2-tensor field that is positive definite at each point. Thus it defines an inner product on T_qM for every $q \in M$ and in every smooth chart (q^i), it has a coordinate representation [1]

$$g = g_{ij}\,\mathrm{d}q^i \otimes \mathrm{d}q^j, \tag{2.10}$$

where g_{ij} is a symmetric positive definite matrix of smooth functions g_{ij}.

Like in Section 2.5 with the symplectic tensor, the metric tensor can also be thought of as a map $\tilde{g}(V) = g(V, \cdot)$ from vector space T_qM to it's dual T_q^*M.

Similarly, $\tilde{g}$ is a linear bijection since $\dim(T_qM) = \dim(T_q^*M)$ and as can be seen from (2.10), $\tilde{g}(V) = 0$ if and only if $V = 0$. Therefore there exists an inverse map $\tilde{g}^{-1} \colon T_q^*M \to T_qM$.

Locally, the inverse $\tilde{g}^{-1}$ is exactly the matrix inverse $\tilde{g}^{-1}(\nu) = g^{-1}(\nu, \cdot)$ and therefore, since g is symmetric positive definite matrix, g^{-1} is symmetric positive matrix and g^{-1} defines a metric on T_q^*M.

Chapter 3

Building the algorithm

3.1 Horizontal forms and lifts induced by projection π

There are few noteworthy vector fields and a horizontal form that arise from the structure of the vector bundle $\pi : E \to M$, which are important exceptions for the upcoming definitions.

Vertical vector fields Y_z on the total space E are defined as vector fields that push forward to the zero vector on the base space M i.e.

$$\pi_{*z} Y_z = 0 \quad \in T_{\pi(z)} M,$$

where $z \in E$ and $Y \in T_z E$.

Similarly *horizontal vector fields* $\tilde{X}_z$ push forward to the tangent space

$$\pi_{*z} \tilde{X}_z = X_{\pi(z)} \quad \in T_{\pi(z)} M. \tag{3.1}$$

Any vector $\tilde{X}_z \in T_z E$ that satisfies the above Equation (3.1), is called a *horizontal lift* of $X \in T_{\pi(z)} M$.

Horizontal forms ω_H are forms that vanish when contracted against a vertical vector field i.e

$$Y \lrcorner \, \omega_H = 0. \tag{3.2}$$

Note that due to the alternating nature of differential forms the above Equation (3.2) is equal to condition

$$\omega_H(V_1, ..., Y_i, ..., V_n) = 0$$

for any $i \in (1, ..., n)$.

In this text a notation $\Omega^k_H(\pi \colon E \to M)$ is used for the space of horizontal forms on total space E.

3.2 Orientation form candidates for the total space

In Section 2, the base manifold M was first expanded into cotangent bundle and then the natural structure of the cotangent bundle made it possible to define the canonical symplectic structure of the cotangent bundle. This was done without any extra information of the form μ_M.

As stated before, the differential form μ_M is a non-vanishing positive top-form, i.e. an orientation form, on the base manifold M.

To take full advantage of the structure that now surrounds the base manifold, the information given by μ_M needs to somehow be encapsulated and "lifted" to the structure defined in the cotangent bundle. Unfortunately, simply using the pull-back $\pi^*\mu_M \in \Omega^n(T^*M)$ is not enough, because an orientation form is needed in the creation of the Hamiltonian function on the cotangent bundle.

The cotangent bundle T^*M is a smooth $2n$-dimensional manifold so the most strait-forward step is to take a smooth n-form $\xi \in \Omega^n(T^*M)$ and wedge it together with pull-back of μ_M to get a top form on the total space that still carries information from the original measure μ_M

$$\mu \equiv \pi^*\mu_M \wedge \xi. \tag{3.3}$$

Surprisingly, the above simple wedge product works remarkably well. The conditions ξ needs to fulfil are not too strict and as it is later seen, ξ actually simplifies into a n-form on the fibres $(T^*M)_q$.

Discussion for what is a practical wise choice for ξ is a subject for a later section 4.1 in this text, but technically it is enough if ξ is smooth, positive in every fibre of T^*M and the integral of ξ over every $(T^*M)_q$ is bounded.

3.2.1 Local expression for μ

For ease of notation, assume local coordinates $(\mathcal{U}, x)$ for $\mathcal{U} \in T^*M$ and the smooth positively oriented frame $(\partial x_1, ..., \partial x_m)$ and the corresponding dual frame $(\mathrm{d}x^1, ..., \mathrm{d}x^m)$ where $m = 2n$.

In this notation the projection $\pi\colon T^*M \to M$ is simply $\pi(x^1, ..., x^n, ..., x^m) = (x^1, ..., x^n)$.

Let I be a rising multi-index of length n, then the elementary n-vectors $(\mathrm{d}x^I)$ form a $\binom{2n}{n}$ dimensional basis for $\Omega^n(T^*M)$ [1], and ξ is locally of the form

$$\begin{aligned} \xi &= \sum_I g_I \mathrm{d}x^I, \quad g_I \in C^\infty(T^*M) \\ &= \sum_{\{I:1\le i_1\le\ldots\le i_n\le m\}} g_{(i_1,\ldots,i_n)}\, \mathrm{d}x^{i_1}\wedge\ldots\wedge\mathrm{d}x^{i_n}. \end{aligned} \tag{3.4}$$

Then μ becomes locally

$$\begin{aligned} \mu &= \pi^*\mu_M \wedge \xi' \\ &= \pi^*(f\,\mathrm{d}x^1\wedge\ldots\wedge\mathrm{d}x^n)\ \wedge \sum_{\{I:1\le i_1\le\ldots\le i_n\le m\}} g_{(i_1,\ldots,i_n)}\,\mathrm{d}x^{i_1}\wedge\ldots\wedge\mathrm{d}x^{i_n} \\ &= \sum_{\{I:1\le i_1\le\ldots\le i_n\le m\}} \pi^*(f\,\mathrm{d}x^1\wedge\ldots\wedge\mathrm{d}x^n)\ \wedge\ g_{(i_1,\ldots,i_n)}\,\mathrm{d}x^{i_1}\wedge\ldots\wedge\mathrm{d}x^{i_n} \\ &= \sum_{\{I:1\le i_1\le\ldots\le i_n\le m\}} (f\circ\pi)\,\pi^*\mathrm{d}x^1\wedge\ldots\wedge\pi^*\mathrm{d}x^n\ \wedge\ g_{(i_1,\ldots,i_n)}\,\mathrm{d}x^{i_1}\wedge\ldots\wedge\mathrm{d}x^{i_n} \\ &= \sum_{\{I:1\le i_1\le\ldots\le i_n\le m\}} (f\circ\pi)g_{(i_1,\ldots,i_n)}\,\pi^*\mathrm{d}x^1\wedge\ldots\wedge\pi^*\mathrm{d}x^n\ \wedge\ \mathrm{d}x^{i_1}\wedge\ldots\wedge\mathrm{d}x^{i_n}. \end{aligned} \tag{3.5}$$

The above expression for μ is still quite messy, but luckily it can still be simplified by first noticing that for any vector V on T^*M and for all covectors $\pi^*\mathrm{d}x^i$, $i=1,..,n$, it holds that

$$\begin{aligned} &\pi^*\mathrm{d}x^1(V) = \mathrm{d}x^1(\pi_*V) \\ &= \pi_*V(x^1) = V(x^1\circ\pi) \\ &= V(x^1) = \mathrm{d}x^1(V). \end{aligned} \tag{3.6}$$

Secondly, due to anticommutativity of wedge product for all repeated indexes it holds that

$$\mathrm{d}x^{i_1}\wedge\mathrm{d}x^{i_1}\wedge\ldots\wedge\mathrm{d}x^{i_k}\,(V_1,..,V_k) = 0.$$

Then in conclusion, the terms in the sum (3.5) for all repeated indexes $i_1 \le n$ vanish and the only surviving non-zero term is

$$(f\circ\pi)g_{(n+1,\ldots,m)}\,\pi^*\mathrm{d}x^1\wedge\ldots\wedge\pi^*\mathrm{d}x^n\wedge\mathrm{d}x^{n+1}\wedge\ldots\wedge\mathrm{d}x^m$$

and therefore the local expression for

$$\mu = \pi^*\mu_M \wedge \xi$$

is

$$\mu = \pi^*(f\,\mathrm{d}x^1 \wedge ... \wedge \mathrm{d}x^n) \;\wedge\; g_{(n+1,...,m)}\,\mathrm{d}x^{n+1} \wedge ... \wedge \mathrm{d}x^m \tag{3.7}$$

or equivalently

$$\mu = (f \circ \pi) g_{(n+1,...,m)}\,\mathrm{d}x^1 \wedge ... \wedge \mathrm{d}x^n \wedge \mathrm{d}x^{n+1} \wedge ... \wedge \mathrm{d}x^m. \tag{3.8}$$

So in the wedge product (3.3), ξ degenerates into a form on the fibre $(T^*M)_q$. Note however, that the function $g_{(n+1,...,m)}$ is still a function from the whole cotangent bundle and can depend on the coordinates $x^i, i \leq n$.

3.2.2 Requirements for ξ

All vectors of the form

$$\begin{aligned} &\sum_{j=k+1}^{n} v^j \partial x_j \\ =&\sum_{j=1}^{k} v^j \frac{\partial}{\partial p_j} \quad \text{(in standard coordinates)} \end{aligned}$$

are vertical vectors by definition. Therefore any forms ξ_1 and ξ_2 differing only by a horizontal form will restrict to the same form on the fibre

$$\begin{aligned} &i_q^*\xi_1\,(V_1, ..., V_k) = i_q^*(\xi_2 + \omega_H)\,(V_1, ..., V_k) \\ &\xi_1(i_{q*}V_1, ..., i_{q*}V_k) = \xi_2(i_{q*}V_1, ..., i_{q*}V_k) + \omega_H(i_{q*}V_1, ..., i_{q*}V_k) \\ &\xi_1(i_{q*}V_1, ..., i_{q*}V_k) = \xi_2(i_{q*}V_1, ..., i_{q*}V_k) \end{aligned}$$

since $i_{q*}V_i$ is a vertical vector.

Motivated by the above, define $\Upsilon^k(\pi\colon T^*M \to M)$ as a quotient space $\Omega^k(T^*M)/\sim$ in which:

1. $\xi_1 \sim \xi_2 \Leftrightarrow \xi_1 - \xi_2 = \omega_H \in \Omega^k_H(\pi\colon T^*M \to M)$.

2. $g_{(n+1,...,m)}$ defined in (3.4) is positive for all points $z \in T^*M$.

3. $\int_{(T^*M)_q} i_q^*\xi = 1$ for all fibres $(T^*M)_q \subset T^*M$, $q \in M$.

Later in this text it is seen that the second condition $g_{(n+1,...,m)} > 0$ is inevitable. It ensures that the basis $(\frac{\partial}{\partial p_1}, ..., \frac{\partial}{\partial p_n})$ from the standard coordinates of the cotangent bundle (2.2) is positively oriented on every fibre T_q^*M with respect to $i_q^*\xi$.

The third condition is there for the random element and for the measure-preserving condition [2] for HMC. In practise, one also needs to be able to sample on every fibre $(T^*M)_q$ from the distribution defined by $g_{(n+1,...,m)}$. Note that in coordinates for this condition it is sufficient if $\int g_{(n+1,...,m)}(q,p)\mathrm{d}p < \infty$ for every $q \in M$ because then $g_{(n+1,...,m)}$ can always be normalised.

Evidently $\Upsilon^k(\pi\colon T^*M \to M)$ is non-empty because

$$\xi = e^{-g_q^{-1}(\varphi,\varphi)}\,\mathrm{d}x^{n+1} \wedge ... \wedge \mathrm{d}x^m, \tag{3.9}$$

where $g_q^{-1}(\varphi,\varphi)$ is the inverse of the Riemannian metric tensor of the base manifold and $\varphi \in T_q^*M$, locally fulfils both later conditions.

Since the base manifold is orientable, any consistent switch of frame introduces a positive Jacobian determinant [1] and hence the form (3.9) is also globally positive in every fibre.

Now confidently the wedge product

$$\mu = \pi^*\mu_M \wedge \xi, \quad \xi \in \Upsilon^k(\pi\colon T^*M \to M) \tag{3.10}$$

is well defined and 'nice' enough to work with.

3.3 A lift of a smooth orientation form

Now that $\mu = \pi^*\mu_M \wedge \xi$ is safely defined, it is time to show that μ is a smooth non-vanishing top form on the cotangent bundle for which the standard coordinate frame is positively oriented. Or in other words: any $\xi \in \Upsilon^n(\pi\colon T^*M \to M)$ lifts any smooth orientation form of the base space to a smooth orientation form on the total space [2].

Let M be a smooth orientable n-manifold with orientation form μ_M on it. Clearly for any $\xi \in \Upsilon(\pi : T^*M \to M)$, μ as a wedge product of two smooth n-forms is a smooth $2n$-form on T^*M [1].

Assume $(\frac{\partial}{\partial q^1}(q), ..., \frac{\partial}{\partial q^n}(q))$ is a positively oriented base on q and $(\mathrm{d}q_1, ..., \mathrm{d}q_n)$ is the corresponding covector base for T_q^*M. This gives access to the standard coordinates (q^i, p_i) of the covector bundle T^*M. Let π be the standard projection $\pi(q,p) = q$.

For any point $q' \in T_q^*M$, $(\frac{\partial}{\partial p_1}(q'), ..., \frac{\partial}{\partial p_n}(q'))$ is a basis on the fibre $(T^*M)_q$. Take any horizontal lift of the base $(\frac{\partial}{\partial q^1}(q), ..., \frac{\partial}{\partial q^n}(q))$ to the point q' to get a base

$$\left(\frac{\tilde{\partial}}{\partial \tilde{q}^1}(q'), ..., \frac{\tilde{\partial}}{\partial \tilde{q}^n}(q'), \frac{\partial}{\partial p_1}(q'), ..., \frac{\partial}{\partial p_n}(q')\right) \tag{3.11}$$

on $T_{q'}(T^*M)$.

Now following the expression in (3.8),

$$\begin{aligned}\mu =&\pi^*\mu_M \wedge \xi \\ =&(f\circ\pi)g_{(p_1,...,p_n)}\,\pi^*\mathrm{d}q^1\wedge ...,\wedge\pi^*\mathrm{d}q^n\;\wedge\;\mathrm{d}p_1\wedge ...\wedge\mathrm{d}p_n\end{aligned} \tag{3.12}$$

note that every vector $\frac{\partial}{\partial p_i}(q')$ is a vertical vector by construction i.e.

$$\pi^*\mathrm{d}q^j\,\left(\frac{\partial}{\partial p_i}(q')\right) = \mathrm{d}q^j\,\pi_*\left(\frac{\partial}{\partial p_i}(q')\right) = \mathrm{d}q^j\,(0) = 0$$

for all $i, j \in (1, ..., n)$. Therefore when μ in (3.12) is evaluated against the base (3.11) on $T_{q'}(T^*M)$,

$$\begin{aligned}&\mu\left(\frac{\tilde{\partial}}{\partial\tilde{q}^1}(q'), ..., \frac{\tilde{\partial}}{\partial\tilde{q}^n}(q'), \frac{\partial}{\partial p_1}(q'), ..., \frac{\partial}{\partial p_n}(q')\right) \\ &= (f\circ\pi)g_{(p_1,...,p_n)}\;\det\begin{pmatrix}\pi^*\mathrm{d}q^1\left(\frac{\tilde{\partial}}{\partial\tilde{q}_1}(q')\right) & \cdots & \pi^*\mathrm{d}q^1\left(\frac{\partial}{\partial p_n}(q')\right) \\ \vdots & \ddots & \vdots \\ \mathrm{d}p^n\left(\frac{\tilde{\partial}}{\partial\tilde{q}_1}(q')\right) & \cdots & \mathrm{d}p^n\left(\frac{\partial}{\partial p_n}(q')\right)\end{pmatrix},\end{aligned} \tag{3.13}$$

the matrix (3.13) is an upper triangular matrix with ones on the diagonal and consequently

$$\begin{aligned}&\mu\left(\frac{\tilde{\partial}}{\partial\tilde{q}^1}(q'), ..., \frac{\tilde{\partial}}{\partial\tilde{q}^n}(q'), \frac{\partial}{\partial p_1}(q'), ..., \frac{\partial}{\partial p_n}(q')\right) \\ &= (f\circ\pi)(q')\cdot g_{(p_1,...,p_n)}(q')\cdot 1 > 0.\end{aligned}$$

Now that finally the information carried by μ_M is transferred to the surrounding total space, the orientation form μ can be used to create a Hamiltonian system on the cotangent bundle.

3.4 Building a Hamiltonian system on the cotangent bundle

The form μ gives orientation to the cotangent bundle T^*M and from (3.13) with similar reasoning it can be seen that the canonical volume form

$$\Omega = \mathrm{d}q^1\wedge ...\wedge\mathrm{d}q^n\wedge\mathrm{d}p_1\wedge ...\wedge\mathrm{d}p_n \tag{3.14}$$

gives the same orientation.

Since the space of top forms $\Omega^{2n}(T^*M)$ is one dimensional [1], there exists such $G\colon T^*M \to \mathbb{R}_+$ that

$$\pi^*\mu_M \wedge \xi = G\,\Omega \tag{3.15}$$

and the function G can be used to define a Hamiltonian vector field X_G i.e.

$$\omega(X_G, \cdot) = \mathrm{d}G,$$

where ω is the canonical symplectic form (2.4)

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

Using the local expressions (3.8) and (3.14) for Ω and $\pi^*\mu \wedge \xi$ gives a local expression for G as

$$\begin{aligned} \pi^*\mu_M \wedge \xi &= G\,\Omega \\ \Rightarrow G &= (f \circ \pi) \cdot g_{(p_1,\ldots,p_n)}. \end{aligned} \tag{3.16}$$

Conventionally however, it is accustomed to use notations

$$\begin{aligned} G &= e^{-H} \\ f \circ \pi &= e^{-V} \\ g_{(p_1,\ldots,p_n)} &= e^{-K}, \end{aligned} \tag{3.17}$$

because then locally the function H becomes neatly the sum of potential V and kinetic energy K [2], [4]

$$\begin{aligned} G &= (f \circ \pi) \cdot g_{(p_1,\ldots,p_n)} \\ H &= V + K, \end{aligned} \tag{3.18}$$

Equation (3.15) becomes

$$\pi^*\mu_M \wedge \xi = e^{-H}\Omega \tag{3.19}$$

with a notable change in the definition of the Hamiltonian vector field:

$$\omega(X_H, \cdot) = \mathrm{d}H. \tag{3.20}$$

These notations are analogous to ones found in physics and are preferred in literature and used through the remainder of this text.

Notice that while $H : T^*M \to \mathbb{R}$ is globally well defined in (3.19), the forms

$$e^{-V}\,\mathrm{d}q^1 \wedge ... \wedge \mathrm{d}q^n = (f \circ \pi)\,\mathrm{d}q^1 \wedge ... \wedge \mathrm{d}q^n$$
$$e^{-K}\,\mathrm{d}p_1 \wedge ... \wedge \mathrm{d}p_n = g_{(p_1,...,p_n)}\,\mathrm{d}p_1 \wedge ... \wedge \mathrm{d}p_n$$

and hence, functions V and K, are only locally defined in (3.16). But since the function H is globally well defined in (3.19), there must be some cancelling effect of chart-dependent terms in the sum $V + K$.

Theorem 3.4.1. *If $(\mathcal{U}, q)$ and $(\tilde{\mathcal{U}}, \tilde{q})$ be two overlapping charts on M then the following identities hold on $\pi^{-1}\Big(\mathcal{U} \cap \tilde{\mathcal{U}}\Big) \in T^*M$.*

$$(f \circ \pi)\,\mathrm{d}q^1 \wedge ... \wedge \mathrm{d}q^n = (f \circ \pi)\,\det\left(\frac{\partial q^j}{\partial \tilde{q}^i}\right)\mathrm{d}\tilde{q}^1 \wedge ... \wedge \mathrm{d}\tilde{q}^n$$

and

$$g_{(p_1,...,p_n)}\,\mathrm{d}p_1 \wedge ... \wedge \mathrm{d}p_n = g_{(p_1,...,p-n)}\,\det\left(\frac{\partial \tilde{q}^i}{\partial q^j}\right)\mathrm{d}\tilde{p}_1 \wedge ... \wedge \mathrm{d}\tilde{p}_n.$$

The first part of the above theorem is proved in [1]. The second part is a long and technical proof by induction and for that reason it can be found in the appendix.

The matrices in the theorem are linear maps between the same two vector bases and one is the inverse of the other. As a result the chart-dependent terms in the sum (3.18) vanish and the sum $V + K$ is chart-independent.

Theorem (3.4.1) also ensures that the local expressions for μ,

$$\mu = e^{-V}\,\mathrm{d}q^1 \wedge ... \wedge \mathrm{d}q^n \wedge e^{-K}\,\mathrm{d}p_1 \wedge ... \wedge \mathrm{d}p_n$$
$$\mu = e^{-(V+K)}\,\mathrm{d}q^1 \wedge ... \wedge \mathrm{d}q^n \wedge \mathrm{d}p_1 \wedge ... \wedge \mathrm{d}p_n$$
$$\mu = e^{-H}\,\mathrm{d}q^1 \wedge ... \wedge \mathrm{d}q^n \wedge \mathrm{d}p_1 \wedge ... \wedge \mathrm{d}p_n$$

are not dependent on the choice of chart, which is very helpful for calculations because then the more concrete local forms of μ can be used interchangeably with the global form $\pi^*\mu_M \wedge \xi$.

3.5 The Markov transition and measure-preserving maps

Now that the tools are ready and in order, the final assembly of HMC can begin.

Let M be a n-dimensional smooth orientable manifold with positively oriented $\mu_M \in \Omega^n(M)$. Let π be the projection $(q,p) \mapsto q$ between the cotangent bundle $p \in T^*M$ and the base manifold $q \in M$. Take any $\xi \in \Upsilon^n(\pi \colon T^*M \to M)$.

The choice of ξ defines the function $g_{(p_1,\ldots,p_n)} = e^{-K} \colon T^*M \to \mathbb{R}_+$ (3.4) and sampling from the fibre $(T^*M)_q$ creates a random lift [2]

$$\begin{aligned}
&\tau \colon M \to T^*M \\
&q \mapsto (q, q'), \text{ where } q' \sim i_q^*\xi \\
&i_q^*\xi = e^{-K(q,\cdot)}\, \mathrm{d}p_1 \wedge \ldots \wedge \mathrm{d}p_n.
\end{aligned}$$

Then create a Hamiltonian vector field X_H

$$\begin{aligned}
&\omega(X_H, \cdot) = \mathrm{d}H \\
&H \colon T^*M \to \mathbb{R} \\
&H = V + K
\end{aligned}$$

and corresponding Hamiltonian flow

$$\theta_t^H \colon T^*M \to T^*M.$$

Compose all these together to get a family of functions

$$\begin{aligned}
&\Theta_{\mathrm{HMC}} \colon M \to M \\
&\Theta_{\mathrm{HMC}} = \pi \circ \theta_t^H \circ \tau,
\end{aligned} \tag{3.21}$$

which is exactly the family of functions discussed in Section 1.4.

One last thing to check is that $Theta_{\mathrm{HMC}}$ is measure-preserving i.e.

$$\mu_M(\Theta_{HMC}^{-1}(U)) = \mu_M(U),$$

for any set $U \subset M$ that is measurable in the sense that

$$\int_U \mu_M \leq 1.$$

Integral over manifold is formally calculated with the help of partition of unity:

Let N be a connected smooth manifold with orientation form μ_M. Assume $\{\psi_\alpha\}$ is a partition of unity subordinate to positively oriented charts $\{(\mathcal{U}_\alpha, \phi_\alpha)\}$ covering M. Then

$$\int_U \mu_M = \sum_\alpha \int_{\phi_\alpha(U_\alpha)} (\phi_\alpha^{-1})^* \psi_\alpha \mu_M$$

is not depended on the choice of charts or on the choice partition of unity [1].

Theorem 3.5.1. *The maps* Θ_{HMC} *defined in* (3.21) *preserve the target measure.*

Before the formal proof there are still qualities in $\pi^*\mu_M \wedge \xi$ that turn out to be useful in the proof of the theorem.

Consider a tensor

$$\pi^*\mu_M \otimes i_q^*\xi \tag{3.22}$$

on $(q, q') \in T^*M$. Motivation for such tensor is not a great leap of faith, because as seen in Section 3.2, form ξ degenerates in the wedge product $\pi^*\mu_M \wedge \xi$ into a form on the fibres.

The above tensor (3.22) is related to the more familiar $\pi^*\mu_M \wedge \xi$ in the following interesting way.

Assume that $A \in T^*M$ is measurable with respect to $(\pi^*\mu_M \wedge \xi)$ and for simplicity, assume that A is contained in a single chart ϕ. Then examine the measure

$$\begin{aligned}
(\pi^*\mu_M \wedge \xi)(A) &= \int_A \pi^*\mu_M \wedge \xi \\
&= \int_{\phi(A)} (\phi^{-1})^*(\pi^*\mu_M \wedge \xi) \\
&= \int_A (\pi^*\mu_M \wedge \xi)\,(\partial q^1, ..., \partial q^n, \partial p_1, ..., \partial p_n) \\
&= \int_A \mathrm{Alt}(\pi^*\mu_M \otimes \xi)(\partial q^1, ..., \partial q^n, \partial p_1, ..., \partial p_n).
\end{aligned}$$

Because ∂p_i are vertical vectors by definition it holds that for any ∂p_i

$$\pi^*\mu_M(V_1, ..., \partial p_i(z), ..., V_n) = 0.$$

Therefore

$$\begin{aligned}
&\int_A \mathrm{Alt}(\pi^*\mu_M \otimes \xi)(\partial q^1, ..., \partial q^n, \partial p_1, ..., \partial p_n) \\
&= \int_A \mathrm{Alt}(\pi^*\mu_M) \otimes \mathrm{Alt}(\xi)\,(\partial q^1, ..., \partial q^n, \partial p_1, ..., \partial p_n) \\
&= \int_A (\pi^*\mu_M \otimes \xi)\,(\partial q^1, ..., \partial q^n, \partial p_1, ..., \partial p_n) \\
&= \int_A (\pi^*\mu_M \otimes i_q^*\xi), \qquad q \in \pi(A)
\end{aligned}$$

or equivalently

$$\int_A \pi^*\mu_M \wedge \xi = \int_A (\pi^*\mu_M \otimes i_q^*\xi) = \int_{\pi(A)} \left(\mu_M(q) \int_{A\,\cap\,(\pi(A))_q} i_q^*\xi \right), \quad (3.23)$$

where $(\pi(A))_q$ is the fibre over $q \in \pi(A)$.

The formula (3.23) states that the measure $\mu(A)$ can be calculated by integrating over each fibre on the projected image of the set A with respect to the measure ξ on the fibres and the measure μ_M on the base manifold. This is an important part in the following proof.

Proof of Theorem 3.5.1. Assume that $U \in M$ is a measurable set, $\{\psi_\alpha\}$ is a partition of unity subordinate to positively oriented charts $\{(\mathcal{U}_\alpha, \phi_\alpha)\}$ covering M and for simplicity, mark E_q as the fibre $(T^*M)_q$ over $q \in M$.

The measure of set U is

$$\mu_M(U) = \int_U \mu_M = \sum_\alpha \int_{U_\alpha \cap U} \psi_\alpha \mu_M.$$

Since

$$\int_{E_q} i_q^*\xi = 1,$$

it holds that

$$\sum_\alpha \int_{U_\alpha \cap U} \psi_\alpha \mu_M = \sum_\alpha \int_{U_\alpha \cap U} \left[\psi_\alpha \mu_M \int_{E_q} i_q^*\xi \right]$$

and by (3.23)

$$\begin{aligned} \sum_\alpha \int_{U_\alpha \cap U} \left[\psi_\alpha \mu_M \int_{E_q} i_q^*\xi \right] &= \int_{(U_\alpha \cap U) \times E_q} \psi_\alpha \pi^*\mu_M \wedge \xi \\ &= \int_{\pi^{-1}(U)} \pi^*\mu_M \wedge \xi. \\ &= \mu(\pi^{-1}(U)). \end{aligned}$$

So the measure of $\mu_M(U)$ is exactly the measure of the whole fibre $\pi^{-1}(U)$ with respect to μ i.e.

$$\mu_M(U) = \mu(\pi^{-1}(U)).$$

Now since flows of smooth vector fields are diffeomorphisms [1], fix $t \in \mathbb{R}$ and consider the measure of the set

$$\int_{\theta^H_{-t}(\pi^{-1}(U))} (\theta^H_{-t})^*(\pi^*\mu_M \wedge \xi),$$

where $\theta^H_{-t} = (\theta^H_t)^{-1}$.

Since H is the Hamiltonian function, then globally by Equation (3.19) and locally by Theorem 3.4.1, the form μ is invariant on the Hamiltonian flow. Therefore

$$\int_{\theta^H_{-t}(\pi^{-1}(U))} (\theta^H_{-t})^*(\pi^*\mu_M \wedge \xi) = \int_B (\pi^*\mu_M \wedge \xi), \quad B = \theta^H_{-t}(\pi^{-1}(U))$$

the measure of $\mu_M(U)$ is connected to the measure of the lifted set $\pi^{-1}U$ that is then pulled-back along the Hamiltonian flow by a string of equalities:

$$\mu_M(U) = \mu(\pi^{-1}(U)) = \mu\left(\theta^H_{-t}\left(\pi^{-1}(U)\right)\right). \tag{3.24}$$

Using (3.23) again gives

$$\int_B (\pi^*\mu_M \wedge \xi) = \int_{\pi(B)} \left(\mu_M \int_{B \,\cap\, (\pi(B))_q} i_q^*\xi\right)$$

and because

$$\int_{B \,\cap\, (\pi(B))_q} i_q^*\xi \leq 1$$

it holds that

$$\int_B (\pi^*\mu_M \wedge \xi) \geq \int_{\pi(B)} \mu_M,$$

which is the final addition to the string (3.24) i.e.

$$\mu_M(U) = \mu(\pi^{-1}(U)) = \mu\left(\theta^H_{-t}\left(\pi^{-1}(U)\right)\right) \geq \mu_M\left(\pi\left(\theta^H_{-t}\left(\pi^{-1}(U)\right)\right)\right)$$

So the measure of the set U with respect to μ_M when lifted to T^*M, mapped along the Hamiltonian flow and finally projected back to M depends on the original measure $\mu_M(U)$ in the following way

$$\begin{aligned} &\mu_M(U) \geq \mu_M\left(\pi\left(\theta^H_{-t}\left(\pi^{-1}(U)\right)\right)\right) \\ \Leftrightarrow\ &\mu_M(U) \geq \mu_M(\Theta^{-1}_{HMC}(U)), \end{aligned} \tag{3.25}$$

which at first looks disappointing because it is not quite what the theorem states. But repeating the steps of the proof from top to bottom to the set $\pi(\theta^H_{-t}(\pi^{-1}(U)))$ while switching the direction of the flow θ^H_{-t} to θ^H_t gives the result as

$$\mu_M(\Theta^{-1}_{HMC}(U)) \geq \mu_M(U) \tag{3.26}$$

and then combining the two Equations (3.25) and (3.26) gives

$$\mu_M(U) = \mu_M(\Theta^{-1}_{HMC}(U)).$$

□

So Θ_{HMC} is a family of measure-preserving maps varying on fibres $q' \sim e^{-K(q,\cdot)}\, \mathrm{d}p_1 \wedge ... \wedge \mathrm{d}p_n$ and possibly on integration time [2] $t \sim \theta^H_t$. Though varying on integration time would require defining a measure on θ^H_t from which to sample from.

Fixing $t \in \mathbb{R}$ in θ^H_t and looking back to the definition of Markov kernel by measure-preserving maps (1.5) (remember that $t \in T$ in (1.5) is different from $t \in \mathbb{R}$ in θ^H_t)

$$\tau(q, A) \equiv \int_T \mathbb{I}_A(t(q))\, g(\mathrm{d}t),$$

$$\mathbb{I}_A(t(q)) \begin{cases} = 1 \text{ if } t(q) \in A \\ = 0 \text{ if } t(q) \notin A, \end{cases}$$

the abstract integral over function space T switches to integral over vector space $(T^*M)_q$ and function $e^{-K(q,\cdot)}$ defines measure on every fibre $(T^*M)_q$. Combining all these gives Markov kernel:

$$\tau(q, A) = \int_{(T^*M)_q} \mathbb{I}_A\left(\pi \circ \theta^H_t(q')\right)\; e^{-K(q,q')}\mathrm{d}q',$$

where $q' \in (T^*M)_q$ and

$$\mathbb{I}_A\left(\pi \circ \theta^H_t(q')\right) \begin{cases} = 1 & \text{if } \pi \circ \theta^H_t(q') \in A \\ = 0 & \text{if } \pi \circ \theta^H_t(q') \notin A. \end{cases}$$

Chapter 4

Further thoughts and practical pointers

4.1 Tying the function K to the metric of the base manifold

Section 3.2 set some loose requirements for the "lifting" form ξ including that the function $g_{(p_1,\ldots,p_n)}$ can but does not have to depend on $q \in M$. This leaves surprisingly free hands to choose the function K to one's liking.

One approach is to tie function $K\colon T^*M \to \mathbb{R}$ to the Riemannian metric of the base manifold and then sample the point q' from some well known distribution on the fibre T_q^*M.

For example choosing K as

$$K(q, q') = \frac{1}{2} g^{-1}(q', q') + \frac{1}{2} \det(g(q)) + \text{constant}, \tag{4.1}$$

where g is the matrix of the Riemannian metric of the base manifold, makes e^{-K} a Gaussian distribution on every fibre T_q^*M [2].

The convergence, and especially the speed of convergence, of Hamiltonian Monte Carlo to the target distribution, is a subject of resent research [8], [9] and it seems to be closely tied to the choice of function K. This is not a surprise considering the local expression for the Hamiltonian vector field introduced in Section 2.5.1.

$$X_H = \sum_{j=1}^{n} \frac{\partial H}{\partial p_j} \frac{\partial}{\partial q^j} - \frac{\partial H}{\partial q^j} \frac{\partial}{\partial p_j}.$$

As can be seen, the movement of the flow relative to the base manifold is controlled by the term $\frac{\partial H}{\partial p_j}$, which in turn is directly related to

$$\frac{\partial g_{(p_1,\ldots,p_n)}}{\partial p_j}. \tag{4.2}$$

So large areas with any significant measure with respect to $i_q^*\xi$ where (4.2) are close to zero should be avoided because sampling to such point would result to only minimal movement on the base manifold. Or in the worst case scenario such areas could make parts of the base manifold almost completely inaccessible to the algorithm. Therefore, for a faster movement with respect to the base manifold, it might be beneficial to choose distribution on the fibres so that it couples high values of the partial derivatives (4.2) with high values of the function $g_{(p_1,\ldots,p_n)}$ itself.

4.2 Few words on symplectic integrators

Symplectic integrators are well researched numeric methods specially designed to solve the Hamiltonian equations (2.9), hence they conserve the symplectic form almost completely even in high dimensions [2].

When working on a general Riemannian manifold, care must be taken when using for example the splitting method of symplectic integration [10]. In splitting method, the Hamiltonian vector field is separated into two terms $X_H = X_K + X_V$. This tempts to separate the Hamiltonian function in the integration simply to $H = V + K$. The method, however, requires that the functions K, V to be chart independent [2] and as shown in Theorem 3.4.1 this is not true for general manifold.

If the function K is chosen like in (4.1) the correct way would be [2]:

$$\begin{aligned} H(q, q') &= H_1 + H_2 \\ H_1 &= \frac{1}{2}\log\left(\det(g(q)) + V(q)\right) \\ H_2 &= \frac{1}{2}g^{-1}(q', q'). \end{aligned}$$

In Euclidean space or on a manifold which has a global chart, it is perfectly fine to separate $H = V + K$ because all the calculations can be done in a single chart and hence the chart dependencies do not come into play.

4.3 Improving and extending HMC

In the practical implementation of Hamiltonian Monte Carlo, there are two important free parameters: integration time and the sample size, that both affect the effectiveness of the algorithm. The problem is that there are no clear general rules of how to choose either of them. Finding good values would then require testing runs of the algorithm that cost computation time.

The No-U-Turn sampler [11] that is in the heart of STAN programming language [2], provides a method that automatically adjusts both integration time and sample size. The No-U-Turn sampler or NUTS, seems to work at least as effectively as HMC with good parameter choices [11].

Another major extensions of HMC are tempering methods [13], [14] and [15] which tackle the tendency of HMC to struggle with distributions that contain multiple isolated modes [14]. Common to these methods is that motivated by thermodynamics, they introduce another variable called temperature that is used to flatten the target distribution while simultaneously lowering the "energy" barriers between different modes [14].

Further future improvements of HMC require solid understanding of the ergodicity of Hamiltonian Monte Carlo [2], [12]. Therefore it might be beneficial to study directly the ergodicity of the Markov chains created by projecting down from the Hamiltonian flow and tying the properties of the chain to the properties of the flow. This could help with optimisation of the algorithm and/or recognising unwanted behaviour before wasting time on computations.

In a recent article [12], the authors manage to set some conditions between geometric ergodicity of the chain and the gradient of potential energy. The proof however, is done with help of an implicit assumption that is not proven in the article. In articles [8] and [9] the convergence results, including geometric ergodicity, are presented without any implicit assumptions. The ergodicity of the chain seems also be achievable with looser conditions if the integration time is allowed to vary [12].

The ideas of Hamiltonian Monte Carlo are also extending beyond finite dimensional spaces to Hilbert spaces [16] and later on, HMC could prove to be useful tool in functional analysis [2].

It is reasonable to expect that Hamiltonian Monte Carlo continues to improve and spread through the scientific community and extend its applications to other branches of mathematics and physics. The path is still though full of theoretical difficulties, in which differential geometry may still have much to give.

Chapter 5

Appendix

5.1 Proof of theorem 3.4.1

Assume smooth n-manifold M and two overlapping charts q and $\tilde{q}$. Let T^*M be the cotangent bundle with standard coordinates (q, p) and $(\tilde{q}, \tilde{p})$ respectively.

The goal is to prove that

$$\mathrm{d}p_1 \wedge ... \wedge \mathrm{d}p_n = \det\left(\frac{\partial \tilde{q}^i}{\partial q^j}\right) \mathrm{d}\tilde{p}_1 \wedge ... \wedge \mathrm{d}\tilde{p}_n,$$

but before starting the actual induction, it is needed to make clear how the change of coordinates in the base space M affect the coordinate components p and especially, how they affect the covectors $\mathrm{d}p$ on the cotangent bundle.

Any point $z \in T^*M$ can be identified as a pair (q, ϕ), where $\phi = p_j \mathrm{d}q^j$ or $(\tilde{q}, \tilde{\phi})$, where $\tilde{\phi} = \tilde{p}_i \mathrm{d}\tilde{q}^i$ depending on the choice of chart. However since both pairs represent the same point in T^*M it must hold that

$$\phi = \tilde{\phi}$$
$$p_i \mathrm{d}q^i = \tilde{p}_i \mathrm{d}\tilde{q}^i$$

and the components p_i and $\tilde{p}_i$ of ϕ in both coordinate frames can be calculated as

$$p_i = \phi(\partial q^i)$$
$$\tilde{p}_i = \phi(\partial \tilde{q}^i).$$

The basis vectors transform from chart q^i to chart $\tilde{q}^i$ as [1]

$$\partial q^i = \frac{\partial \tilde{q}^j}{\partial q^i} \partial \tilde{q}^j.$$

Then the components of ϕ in terms of $\tilde{q}^i$ are

$$p_i = \phi\left(\frac{\partial \tilde{q}^j}{\partial q^i}\partial \tilde{q}^j\right) = \frac{\partial \tilde{q}^j}{\partial q^i}\phi(\partial \tilde{q}^j) = \frac{\partial \tilde{q}^j}{\partial q^i}\tilde{p}_j.$$

The above is change of coordinates for a general ϕ, but what is needed in this proof is change of coordinates formula for $\mathrm{d}p_i$.

Write $\phi = \mathrm{d}p_i$.

Then the components p_i can be calculated

$$\begin{cases} p_1 & = \mathrm{d}p_i(\partial q^1) = 0 \\ & \vdots \\ p_i & = \mathrm{d}p_i(\partial q^i) = 1 \\ & \vdots \\ p_n & = \mathrm{d}p_i(\partial q^n) = 0 \end{cases}$$

$$\implies p_j = \mathrm{d}p_i(\partial q^j) = \delta^i_j.$$

So p_i in terms of $\tilde{q}^i$ is

$$p_i = \mathrm{d}p_i\left(\frac{\partial \tilde{q}^j}{\partial q^i}\partial \tilde{q}^j\right) = \left(\frac{\partial \tilde{q}^j}{\partial q^i}\right)\mathrm{d}p_i(\partial \tilde{q}^j) = \left(\frac{\partial \tilde{q}^j}{\partial q^i}\right)\tilde{p}_j$$

and finally the expression for $\mathrm{d}p_i$ in terms of $\tilde{q}^i$ is

$$\begin{aligned} \mathrm{d}p_i &= \mathrm{d}\left(\left(\frac{\partial \tilde{q}^j}{\partial q^i}\right)\tilde{p}_j\right) \\ &= \left(\frac{\partial \tilde{q}^1}{\partial q^i}\right)\mathrm{d}(\tilde{p}_1) + \ldots + \left(\frac{\partial \tilde{q}^n}{\partial q^i}\right)\mathrm{d}(\tilde{p}_n) \\ &= \left(\frac{\partial \tilde{q}^j}{\partial q^i}\right)\mathrm{d}\tilde{p}_j. \end{aligned}$$

Proof by induction. Base case: Assume that $n = 2$.

Writing $\mathrm{d}p_1 \wedge \mathrm{d}p_2$ in terms of $\tilde{q}^i$ creates expression

$$\begin{aligned} &\mathrm{d}p_1 \wedge \mathrm{d}p_2 \\ &= \left(\frac{\partial \tilde{q}^1}{\partial q^1}\mathrm{d}\tilde{p}_1 + \frac{\partial \tilde{q}^2}{\partial q^1}\mathrm{d}\tilde{p}_2\right) \wedge \left(\frac{\partial \tilde{q}^1}{\partial q^2}\mathrm{d}\tilde{p}_1 + \frac{\partial \tilde{q}^2}{\partial q^2}\mathrm{d}\tilde{p}_2\right). \end{aligned} \tag{5.1}$$

Since the wedge product is bilinear it holds that

$$\left(\frac{\partial \tilde{q}^1}{\partial q^1}\mathrm{d}\tilde{p}_1 + \frac{\partial \tilde{q}^2}{\partial q^1}\mathrm{d}\tilde{p}_2\right) \wedge \left(\frac{\partial \tilde{q}^1}{\partial q^2}\mathrm{d}\tilde{p}_1 + \frac{\partial \tilde{q}^2}{\partial q^2}\mathrm{d}\tilde{p}_2\right)$$

$$= \left(\frac{\partial \tilde{q}^1}{\partial q^1}\mathrm{d}\tilde{p}_1 + \frac{\partial \tilde{q}^2}{\partial q^1}\mathrm{d}\tilde{p}_2\right) \wedge \frac{\partial \tilde{q}^1}{\partial q^2}\mathrm{d}\tilde{p}_1 + \left(\frac{\partial \tilde{q}^1}{\partial q^1}\mathrm{d}\tilde{p}_1 + \frac{\partial \tilde{q}^2}{\partial q^1}\mathrm{d}\tilde{p}_2\right) \wedge \frac{\partial \tilde{q}^2}{\partial q^2}\mathrm{d}\tilde{p}_2$$

$$= \frac{\partial \tilde{q}^1}{\partial q^1}\mathrm{d}\tilde{p}_1 \wedge \frac{\partial \tilde{q}^1}{\partial q^2}\mathrm{d}\tilde{p}_1 + \frac{\partial \tilde{q}^2}{\partial q^1}\mathrm{d}\tilde{p}_2 \wedge \frac{\partial \tilde{q}^1}{\partial q^2}\mathrm{d}\tilde{p}_1 + \frac{\partial \tilde{q}^1}{\partial q^1}\mathrm{d}\tilde{p}_1 \wedge \frac{\partial \tilde{q}^2}{\partial q^2}\mathrm{d}\tilde{p}_2 + \frac{\partial \tilde{q}^2}{\partial q^1}\mathrm{d}\tilde{p}_2 \wedge \frac{\partial \tilde{q}^2}{\partial q^2}\mathrm{d}\tilde{p}_2.$$

Due to the anticommutativity of wedge product, the terms marked in orange are zeros and the remaining terms can be rearranged to make use of bilinearity again:

$$\begin{aligned}
&\frac{\partial \tilde{q}^2}{\partial q^1}\mathrm{d}\tilde{p}_2 \wedge \frac{\partial \tilde{q}^1}{\partial q^2}\mathrm{d}\tilde{p}_1 + \frac{\partial \tilde{q}^1}{\partial q^1}\mathrm{d}\tilde{p}_1 \wedge \frac{\partial \tilde{q}^2}{\partial q^2}\mathrm{d}\tilde{p}_2 \\
&= \frac{\partial \tilde{q}^1}{\partial q^1}\mathrm{d}\tilde{p}_1 \wedge \frac{\partial \tilde{q}^2}{\partial q^2}\mathrm{d}\tilde{p}_2 - \frac{\partial \tilde{q}^1}{\partial q^2}\mathrm{d}\tilde{p}_1 \wedge \frac{\partial \tilde{q}^2}{\partial q^1}\mathrm{d}\tilde{p}_2 \\
&= \frac{\partial \tilde{q}^1}{\partial q^1}\frac{\partial \tilde{q}^2}{\partial q^2}\mathrm{d}\tilde{p}_1 \wedge \mathrm{d}\tilde{p}_2 - \frac{\partial \tilde{q}^1}{\partial q^2}\frac{\partial \tilde{q}^2}{\partial q^1}\mathrm{d}\tilde{p}_1 \wedge \mathrm{d}\tilde{p}_2 \\
&= \left(\frac{\partial \tilde{q}^1}{\partial q^1}\frac{\partial \tilde{q}^2}{\partial q^2} - \frac{\partial \tilde{q}^1}{\partial q^2}\frac{\partial \tilde{q}^2}{\partial q^1}\right)\mathrm{d}\tilde{p}_1 \wedge \mathrm{d}\tilde{p}_2 \\
&= \det\left(\frac{\partial \tilde{q}^i}{\partial q^j}\right)\mathrm{d}\tilde{p}_1 \wedge \mathrm{d}\tilde{p}_2.
\end{aligned}$$

Therefore

$$\mathrm{d}p_1 \wedge \mathrm{d}p_2 = \det\left(\frac{\partial \tilde{q}^i}{\partial q^j}\right)\mathrm{d}\tilde{p}_1 \wedge \mathrm{d}\tilde{p}_2 \tag{5.2}$$

and the theorem holds for $n = 2$.

Induction step: Assume that for some n, it holds that

$$\mathrm{d}p_1 \wedge ... \wedge \mathrm{d}p_n = \det\left(\frac{\partial \tilde{q}^j}{\partial q^i}\right)\mathrm{d}\tilde{p}_1 \wedge ... \wedge \mathrm{d}\tilde{p}_n.$$

Like in (5.1), writing

$$\mathrm{d}p_1 \wedge ... \wedge \mathrm{d}p_n \wedge \mathrm{d}p_{n+1}$$

in terms of $\tilde{q}^i$ gives

$$
\begin{aligned}
&\mathrm{d}p_1 \wedge \dots \wedge \mathrm{d}p_n \wedge \mathrm{d}p_{n+1} \\
=&\sum_{j=1}^{n+1} \frac{\partial \tilde{q}^j}{\partial q^1}\mathrm{d}\tilde{p}_j \wedge \dots \wedge \sum_{j=1}^{n+1} \frac{\partial \tilde{q}^j}{\partial q^n}\mathrm{d}\tilde{p}_j \wedge \sum_{j=1}^{n+1} \frac{\partial \tilde{q}^j}{\partial q^{n+1}}\mathrm{d}\tilde{p}_j \qquad (5.3) \\
=&\left(\frac{\partial \tilde{q}^1}{\partial q^1}\mathrm{d}\tilde{p}_1 + \dots + \frac{\partial \tilde{q}^n}{\partial q^1}\mathrm{d}\tilde{p}_n + \frac{\partial \tilde{q}^{n+1}}{\partial q^1}\mathrm{d}\tilde{p}_{n+1}\right) \\
&\wedge \dots \wedge \\
&\wedge \left(\frac{\partial \tilde{q}^1}{\partial q^n}\mathrm{d}\tilde{p}_1 + \dots + \frac{\partial \tilde{q}^n}{\partial q^n}\mathrm{d}\tilde{p}_n + \frac{\partial \tilde{q}^{n+1}}{\partial q^n}\mathrm{d}\tilde{p}_{n+1}\right) \\
&\wedge \left(\frac{\partial \tilde{q}^1}{\partial q^{n+1}}\mathrm{d}\tilde{p}_1 + \dots + \frac{\partial \tilde{q}^n}{\partial q^{n+1}}\mathrm{d}\tilde{p}_n + \frac{\partial \tilde{q}^{n+1}}{\partial q^{n+1}}\mathrm{d}\tilde{p}_{n+1}\right).
\end{aligned}
$$

Calculating the first wedge product in (5.3) gives

$$
\begin{aligned}
&\left(\frac{\partial \tilde{q}^1}{\partial q^1}\mathrm{d}\tilde{p}_1 + \dots + \frac{\partial \tilde{q}^n}{\partial q^1}\mathrm{d}\tilde{p}_n + \frac{\partial \tilde{q}^{n+1}}{\partial q^1}\mathrm{d}\tilde{p}_{n+1}\right) \\
&\wedge \left(\frac{\partial \tilde{q}^1}{\partial q^2}\mathrm{d}\tilde{p}_1 + \dots + \frac{\partial \tilde{q}^n}{\partial q^2}\mathrm{d}\tilde{p}_n + \frac{\partial \tilde{q}^{n+1}}{\partial q^2}\mathrm{d}\tilde{p}_{n+1}\right) \wedge \dots \\
=&\Bigg[\left(\frac{\partial \tilde{q}^1}{\partial q^1}\mathrm{d}\tilde{p}_1 + \dots + \frac{\partial \tilde{q}^n}{\partial q^1}\mathrm{d}\tilde{p}_n\right) \wedge \left(\frac{\partial \tilde{q}^1}{\partial q^2}\mathrm{d}\tilde{p}_1 + \dots + \frac{\partial \tilde{q}^n}{\partial q^2}\mathrm{d}\tilde{p}_n\right) \\
&+ \left(\frac{\partial \tilde{q}^1}{\partial q^1}\mathrm{d}\tilde{p}_1 + \dots + \frac{\partial \tilde{q}^n}{\partial q^1}\mathrm{d}\tilde{p}_n\right) \wedge \frac{\partial \tilde{q}^{n+1}}{\partial q^2}\mathrm{d}\tilde{p}_{n+1} \\
&+ \frac{\partial \tilde{q}^{n+1}}{\partial q^1}\mathrm{d}\tilde{p}_{n+1} \wedge \left(\frac{\partial \tilde{q}^1}{\partial q^2}\mathrm{d}\tilde{p}_1 + \dots + \frac{\partial \tilde{q}^n}{\partial q^2}\mathrm{d}\tilde{p}_n\right) \\
&+ \frac{\partial \tilde{q}^{n+1}}{\partial q^1}\mathrm{d}\tilde{p}_{n+1} \wedge \frac{\partial \tilde{q}^{n+1}}{\partial q^2}\mathrm{d}\tilde{p}_{n+1}\Bigg] \wedge \dots
\end{aligned}
$$

where the last term in the sum marked in orange is zero because of anti-commutativity of wedge product.

Notice how the $\mathrm{d}\tilde{p}_{n+1}$ terms move around. To help visualise this, consider product of sums

$$
\begin{aligned}
&(A_1 + B_1 + C_1) \cdot (A_2 + B_2 + C_2) \\
=&(A_1 + B_1 + C_1) \cdot A_2 + (A_1 + B_1 + C_1) \cdot B_2 + (A_1 + B_1 + C_1) \cdot C_2 \\
=&(A_1 + B_1) \cdot A_2 + (A_1 + B_1) \cdot B_2 + (A_1 + B_1) \cdot C_2 \;+\; C_1 \cdot A_2 + C_1 \cdot B_2 + C_1 \cdot C_2 \\
=&(A_1 + B_1) \cdot (A_2 + B_2) \\
&+ C_1 \cdot (A_2 + B_2) \\
&+ (A_1 + B_1) \cdot C_2 \\
&+ C_1 \cdot C_2,
\end{aligned}
$$

where the lower indexes represent some coefficients. Then adding another sum in the product gives

$$
\begin{aligned}
&(A_1 + B_1 + C_1) \cdot (A_2 + B_2 + C_2) \cdot (A_3 + B_3 + C_3) \\
=&\Big[(A_1 + B_1) \cdot (A_2 + B_2) \\
&+ C_1 \cdot (A_2 + B_2) \\
&+ (A_1 + B_1) \cdot C_2 \\
&+ C_1 \cdot C_2\Big] \\
&\cdot (A_3 + B_3 + C_3) \\
=&(A_1 + B_1) \cdot (A_2 + B_2) \cdot (A_3 + B_3) \\
&+ C_1 \cdot (A_2 + B_2) \cdot (A_3 + B_3) \\
&+ (A_1 + B_1) \cdot C_2 \cdot (A_3 + B_3) \\
&+ (A_1 + B_1) \cdot (A_2 + B_2) \cdot C_3 \\
&+ C_1 \cdot (A_2 + B_2) \cdot C_3 \;+\; (A_1 + B_1) \cdot C_2 \cdot C_3 \;+\; C_1 \cdot C_2 \cdot C_3.
\end{aligned}
\tag{5.4}
$$

Imagine that instead of product '$\cdot$', there is anticommutative wedge product and instead of A_i, B_i, C_i, there is $\frac{\partial \tilde{q}^1}{\partial q^i} \mathrm{d}\tilde{p}_1$, $\frac{\partial \tilde{q}^n}{\partial q^i} \mathrm{d}\tilde{p}_n$ and $\frac{\partial \tilde{q}^{n+1}}{\partial q^i} \mathrm{d}\tilde{p}_{n+1}$ respectively. Then the final line, having the products containing multiple C-terms will vanish due to anticommutativity.

Now switching back to (5.3) and continuing the calculations to the third wedge product $\sum_{j=1}^{n+1} \frac{\partial \tilde{q}^j}{\partial q^3} \mathrm{d}\tilde{q}^j$:

$$\Bigg[\left(\frac{\partial \tilde{q}^1}{\partial q^1} \mathrm{d}\tilde{p}_1 + ... + \frac{\partial \tilde{q}^n}{\partial q^1} \mathrm{d}\tilde{p}_n \right) \wedge \left(\frac{\partial \tilde{q}^1}{\partial q^2} \mathrm{d}\tilde{p}_1 + ... + \frac{\partial \tilde{q}^n}{\partial q^2} \mathrm{d}\tilde{p}_n \right)$$
$$+ \left(\frac{\partial \tilde{q}^1}{\partial q^1} \mathrm{d}\tilde{p}_1 + ... + \frac{\partial \tilde{q}^n}{\partial q^1} \mathrm{d}\tilde{p}_n \right) \wedge \frac{\partial \tilde{q}^{n+1}}{\partial q^2} \mathrm{d}\tilde{p}_{n+1}$$
$$+ \frac{\partial \tilde{q}^{n+1}}{\partial q^1} \mathrm{d}\tilde{p}_{n+1} \wedge \left(\frac{\partial \tilde{q}^1}{\partial q^2} \mathrm{d}\tilde{p}_1 + ... + \frac{\partial \tilde{q}^n}{\partial q^2} \mathrm{d}\tilde{p}_n \right) \Bigg]$$
$$\wedge \left(\frac{\partial \tilde{q}^1}{\partial q^3} \mathrm{d}\tilde{p}_1 + ... + \frac{\partial \tilde{q}^n}{\partial q^3} \mathrm{d}\tilde{p}_n + \frac{\partial \tilde{q}^{n+1}}{\partial q^3} \mathrm{d}\tilde{p}_{n+1} \right) \wedge ...$$
$$= \Bigg[\left(\frac{\partial \tilde{q}^1}{\partial q^1} \mathrm{d}\tilde{p}_1 + ... + \frac{\partial \tilde{q}^n}{\partial q^1} \mathrm{d}\tilde{p}_n \right) \wedge \left(\frac{\partial \tilde{q}^1}{\partial q^2} \mathrm{d}\tilde{p}_1 + ... + \frac{\partial \tilde{q}^n}{\partial q^2} \mathrm{d}\tilde{p}_n \right) \wedge \left(\frac{\partial \tilde{q}^1}{\partial q^3} \mathrm{d}\tilde{p}_1 + ... + \frac{\partial \tilde{q}^n}{\partial q^3} \mathrm{d}\tilde{p}_n \right)$$
$$+ \left(\frac{\partial \tilde{q}^1}{\partial q^1} \mathrm{d}\tilde{p}_1 + ... + \frac{\partial \tilde{q}^n}{\partial q^1} \mathrm{d}\tilde{p}_n \right) \wedge \left(\frac{\partial \tilde{q}^1}{\partial q^2} \mathrm{d}\tilde{p}_1 + ... + \frac{\partial \tilde{q}^n}{\partial q^2} \mathrm{d}\tilde{p}_n \right) \wedge \frac{\partial \tilde{q}^{n+1}}{\partial q^3} \mathrm{d}\tilde{p}_{n+1}$$
$$+ \left(\frac{\partial \tilde{q}^1}{\partial q^1} \mathrm{d}\tilde{p}_1 + ... + \frac{\partial \tilde{q}^n}{\partial q^1} \mathrm{d}\tilde{p}_n \right) \wedge \frac{\partial \tilde{q}^{n+1}}{\partial q^2} \mathrm{d}\tilde{p}_{n+1} \wedge \left(\frac{\partial \tilde{q}^1}{\partial q^3} \mathrm{d}\tilde{p}_1 + ... + \frac{\partial \tilde{q}^n}{\partial q^3} \mathrm{d}\tilde{p}_n \right)$$
$$+ \frac{\partial \tilde{q}^{n+1}}{\partial q^1} \mathrm{d}\tilde{p}_{n+1} \wedge \left(\frac{\partial \tilde{q}^1}{\partial q^2} \mathrm{d}\tilde{p}_1 + ... + \frac{\partial \tilde{q}^n}{\partial q^2} \mathrm{d}\tilde{p}_n \right) \wedge \left(\frac{\partial \tilde{q}^1}{\partial q^3} \mathrm{d}\tilde{p}_1 + ... + \frac{\partial \tilde{q}^n}{\partial q^3} \mathrm{d}\tilde{p}_n \right) \Bigg] \wedge ...,$$

where the wedge products containing multiple terms $\mathrm{d}\tilde{p}_{n+1}$ like in (5.4), are omitted because they would be zeros again due to anticommutativity.

The sum notation gives less messy expression for the above third wedge product as

$$\Bigg[\left(\sum_{j=1}^{n} \frac{\partial \tilde{q}^j}{\partial q^1} \mathrm{d}\tilde{p}_j \right) \wedge \left(\sum_{j=1}^{n} \frac{\partial \tilde{q}^j}{\partial q^2} \mathrm{d}\tilde{p}_j \right) \wedge \left(\sum_{j=1}^{n} \frac{\partial \tilde{q}^j}{\partial q^3} \mathrm{d}\tilde{p}_j \right) \tag{5.5}$$
$$+ \frac{\partial \tilde{q}^{n+1}}{\partial q^1} \mathrm{d}\tilde{p}_{n+1} \wedge \left(\sum_{j=1}^{n} \frac{\partial \tilde{q}^j}{\partial q^2} \mathrm{d}\tilde{p}_j \right) \wedge \left(\sum_{j=1}^{n} \frac{\partial \tilde{q}^j}{\partial q^3} \mathrm{d}\tilde{p}_j \right) \tag{5.6}$$
$$+ \left(\sum_{j=1}^{n} \frac{\partial \tilde{q}^j}{\partial q^1} \mathrm{d}\tilde{p}_j \right) \wedge \frac{\partial \tilde{q}^{n+1}}{\partial q^2} \mathrm{d}\tilde{p}_{n+1} \wedge \left(\sum_{j=1}^{n} \frac{\partial \tilde{q}^j}{\partial q^3} \mathrm{d}\tilde{p}_j \right) \tag{5.7}$$
$$+ \left(\sum_{j=1}^{n} \frac{\partial \tilde{q}^j}{\partial q^1} \mathrm{d}\tilde{p}_j \right) \wedge \left(\sum_{j=1}^{n} \frac{\partial \tilde{q}^j}{\partial q^2} \mathrm{d}\tilde{p}_j \right) \wedge \frac{\partial \tilde{q}^{n+1}}{\partial q^3} \mathrm{d}\tilde{p}_{n+1} \Bigg] \wedge ... \tag{5.8}$$

Then following the above pattern and continuing the calculations through all the wedge products in (5.3), gives a expression:

$$
\begin{aligned}
&\left(\sum_{j=1}^{n}\frac{\partial\tilde{q}^j}{\partial q^1}\mathrm{d}\tilde{p}_j\right)\wedge\ldots\wedge\left(\sum_{j=1}^{n}\frac{\partial\tilde{q}^j}{\partial q^n}\mathrm{d}\tilde{p}_j\right)\wedge\left(\sum_{j=1}^{n}\frac{\partial\tilde{q}^j}{\partial q^{n+1}}\mathrm{d}\tilde{p}_j\right)\\
&+\frac{\partial\tilde{q}^{n+1}}{\partial q^1}\mathrm{d}\tilde{p}_{n+1}\wedge\left(\sum_{j=1}^{n}\frac{\partial\tilde{q}^j}{\partial q^2}\mathrm{d}\tilde{p}_j\right)\wedge\ldots\wedge\left(\sum_{j=1}^{n}\frac{\partial\tilde{q}^j}{\partial q^{n+1}}\mathrm{d}\tilde{p}_j\right)\\
&+\left(\sum_{j=1}^{n}\frac{\partial\tilde{q}^j}{\partial q^1}\mathrm{d}\tilde{p}_j\right)\wedge\frac{\partial\tilde{q}^{n+1}}{\partial q^2}\mathrm{d}\tilde{p}_{n+1}\wedge\ldots\wedge\left(\sum_{j=1}^{n}\frac{\partial\tilde{q}^j}{\partial q^{n+1}}\mathrm{d}\tilde{p}_j\right)\\
&+\ldots+\\
&+\left(\sum_{j=1}^{n}\frac{\partial\tilde{q}^j}{\partial q^1}\mathrm{d}\tilde{p}_j\right)\wedge\ldots\wedge\frac{\partial\tilde{q}^{n+1}}{\partial q^i}\mathrm{d}\tilde{p}_{n+1}\wedge\ldots\wedge\left(\sum_{j=1}^{n}\frac{\partial\tilde{q}^j}{\partial q^{n+1}}\mathrm{d}\tilde{p}_j\right)\\
&+\ldots+\\
&+\left(\sum_{j=1}^{n}\frac{\partial\tilde{q}^j}{\partial q^1}\mathrm{d}\tilde{p}_j\right)\wedge\ldots\wedge\left(\sum_{j=1}^{n}\frac{\partial\tilde{q}^j}{\partial q^n}\mathrm{d}\tilde{p}_j\right)\wedge\frac{\partial\tilde{q}^{n+1}}{\partial q^{n+1}}\mathrm{d}\tilde{p}_{n+1}.
\end{aligned}
\tag{5.9}
$$

In (5.9), the first term marked in orange is a $(n+1)$th wedge product of n different 1-forms, which makes it zero. Another way of seeing this, is by making use of the induction assumption.

$$
\begin{aligned}
&\det\left(\frac{\partial\tilde{q}^j}{\partial q^i}\right)\mathrm{d}\tilde{p}_1\wedge\ldots\wedge\mathrm{d}\tilde{p}_n\\
&=\mathrm{d}p_1\wedge\ldots\wedge\mathrm{d}p_n\\
&=\left(\sum_{j=1}^{n}\frac{\partial\tilde{q}^j}{\partial q^1}\mathrm{d}\tilde{p}_j\right)\wedge\ldots\wedge\left(\sum_{j=1}^{n}\frac{\partial\tilde{q}^j}{\partial q^n}\mathrm{d}\tilde{p}_j\right),
\end{aligned}
$$

where the last row is just $\mathrm{d}p_1\wedge\ldots\wedge\mathrm{d}p_n$ written in terms of $\tilde{q}^i$. Now the orange term in (5.9) equals to

$$\begin{aligned}
&\left(\sum_{j=1}^{n}\frac{\partial \tilde{q}^j}{\partial q^1}\mathrm{d}\tilde{p}_j\right)\wedge\ldots\wedge\left(\sum_{j=1}^{n}\frac{\partial \tilde{q}^j}{\partial q^n}\mathrm{d}\tilde{p}_j\right)\wedge\left(\sum_{j=1}^{n}\frac{\partial \tilde{q}^j}{\partial q^{n+1}}\mathrm{d}\tilde{p}_j\right)\\
&=\det\left(\frac{\partial \tilde{q}^j}{\partial q^i}\right)\mathrm{d}\tilde{p}_1\wedge\ldots\wedge\mathrm{d}\tilde{p}_n\wedge\left(\sum_{j=1}^{n}\frac{\partial \tilde{q}^j}{\partial q^{n+1}}\mathrm{d}\tilde{p}_j\right)\\
&=\det\left(\frac{\partial \tilde{q}^j}{\partial q^i}\right)\mathrm{d}\tilde{p}_1\wedge\ldots\wedge\mathrm{d}\tilde{p}_n\wedge\frac{\partial \tilde{q}^1}{\partial q^{n+1}}\mathrm{d}\tilde{p}_1\\
&\quad+\det\left(\frac{\partial \tilde{q}^j}{\partial q^i}\right)\mathrm{d}\tilde{p}_1\wedge\ldots\wedge\mathrm{d}\tilde{p}_n\wedge\frac{\partial \tilde{q}^2}{\partial q^{n+1}}\mathrm{d}\tilde{p}_2\\
&\quad+\ldots+\\
&\quad+\det\left(\frac{\partial \tilde{q}^j}{\partial q^i}\right)\mathrm{d}\tilde{p}_1\wedge\ldots\wedge\mathrm{d}\tilde{p}_n\wedge\frac{\partial \tilde{q}^n}{\partial q^{n+1}}\mathrm{d}\tilde{p}_n,
\end{aligned}$$

where every single wedge product

$$\det\left(\frac{\partial \tilde{q}^j}{\partial q^i}\right)\mathrm{d}\tilde{p}_1\wedge\ldots\wedge\mathrm{d}\tilde{p}_i\wedge\ldots\wedge\mathrm{d}\tilde{p}_n\wedge\frac{\partial \tilde{q}^i}{\partial q^{n+1}}\mathrm{d}\tilde{p}_i=0$$

due to anticommutativity, because the covectors $\mathrm{d}\tilde{p}_i$ appear twice for every $i=1,\ldots,n$.

Returning back to (5.9) and rearranging the remaining terms in the wedge product gives the following sum of $n+1$ forms

$$\begin{aligned}
&\left(\sum_{j=1}^{n}\frac{\partial \tilde{q}^j}{\partial q^2}\mathrm{d}\tilde{p}_j\right)\wedge\ldots\wedge\left(\sum_{j=1}^{n}\frac{\partial \tilde{q}^j}{\partial q^{n+1}}\mathrm{d}\tilde{p}_j\right)\wedge(-1)^{(n+1)-1}\cdot\frac{\partial \tilde{q}^{n+1}}{\partial q^1}\mathrm{d}\tilde{p}_{n+1}\\
&+\left(\sum_{j=1}^{n}\frac{\partial \tilde{q}^j}{\partial q^1}\mathrm{d}\tilde{p}_j\right)\wedge\ldots\wedge\left(\sum_{j=1}^{n}\frac{\partial \tilde{q}^j}{\partial q^{n+1}}\mathrm{d}\tilde{p}_j\right)\wedge(-1)^{(n+1)-2}\cdot\frac{\partial \tilde{q}^{n+1}}{\partial q^2}\mathrm{d}\tilde{p}_{n+1}\\
&+\ldots+\\
&+\left(\sum_{j=1}^{n}\frac{\partial \tilde{q}^j}{\partial q^1}\mathrm{d}\tilde{p}_j\right)\wedge\ldots\wedge\left(\sum_{j=1}^{n}\frac{\partial \tilde{q}^j}{\partial q^{n+1}}\mathrm{d}\tilde{p}_j\right)\wedge(-1)^{(n+1)-i}\cdot\frac{\partial \tilde{q}^{n+1}}{\partial q^i}\mathrm{d}\tilde{p}_{n+1}\\
&+\ldots+\\
&+\left(\sum_{j=1}^{n}\frac{\partial \tilde{q}^j}{\partial q^1}\mathrm{d}\tilde{p}_j\right)\wedge\ldots\wedge\left(\sum_{j=1}^{n}\frac{\partial \tilde{q}^j}{\partial q^{n+1}}\mathrm{d}\tilde{p}_j\right)\wedge(-1)\cdot\frac{\partial \tilde{q}^{n+1}}{\partial q^n}\mathrm{d}\tilde{p}_{n+1}\\
+&\left(\sum_{j=1}^{n}\frac{\partial \tilde{q}^j}{\partial q^1}\mathrm{d}\tilde{p}_j\right)\wedge\ldots\wedge\left(\sum_{j=1}^{n}\frac{\partial \tilde{q}^j}{\partial q^n}\mathrm{d}\tilde{p}_j\right)\wedge\frac{\partial \tilde{q}^{n+1}}{\partial q^{n+1}}\mathrm{d}\tilde{p}_{n+1}.
\end{aligned}$$

The next step is to again make use of the assumption of the induction step:

$$
\begin{aligned}
&\det\left(\frac{\partial \tilde{q}^j}{\partial q^i}\right) \mathrm{d}\tilde{p}_1 \wedge ... \wedge \mathrm{d}\tilde{p}_n \\
&= \mathrm{d}p_1 \wedge ... \wedge \mathrm{d}p_n \\
&= \left(\sum_{j=1}^{n} \frac{\partial \tilde{q}^j}{\partial q^1} \mathrm{d}\tilde{p}_j\right) \wedge ... \wedge \left(\sum_{j=1}^{n} \frac{\partial \tilde{q}^j}{\partial q^n} \mathrm{d}\tilde{p}_j\right).
\end{aligned}
$$

Note however that in this form, the assumption works only on the last row of (5.9)

$$
\begin{aligned}
&\left(\sum_{j=1}^{n} \frac{\partial \tilde{q}^j}{\partial q^1} \mathrm{d}\tilde{p}_j\right) \wedge ... \wedge \left(\sum_{j=1}^{n} \frac{\partial \tilde{q}^j}{\partial q^n} \mathrm{d}\tilde{p}_j\right) \wedge \frac{\partial \tilde{q}^{n+1}}{\partial q^{n+1}} \mathrm{d}\tilde{p}_{n+1} \\
&= \det\left(\frac{\partial \tilde{q}^j}{\partial q^i}\right) \mathrm{d}\tilde{p}_1 \wedge ... \wedge \mathrm{d}\tilde{p}_n \wedge \frac{\partial \tilde{q}^{n+1}}{\partial q^{n+1}} \mathrm{d}\tilde{p}_{n+1},
\end{aligned}
\tag{5.10}
$$

since in every other row

$$
\left(\sum_{j=1}^{n} \frac{\partial \tilde{q}^j}{\partial q^1} \mathrm{d}\tilde{p}_j\right) \wedge ... \wedge \left(\sum_{j=1}^{n} \frac{\partial \tilde{q}^j}{\partial q^{n+1}} \mathrm{d}\tilde{p}_j\right) \wedge (-1)^{(n+1)-i} \frac{\partial \tilde{q}^{n+1}}{\partial q^i} \mathrm{d}\tilde{p}_{n+1}
$$

the sum

$$
\left(\sum_{j=1}^{n} \frac{\partial \tilde{q}^j}{\partial q^i} \mathrm{d}\tilde{p}_j\right)
$$

is missing from the wedge product. But luckily the term

$$
\left(\sum_{j=1}^{n} \frac{\partial \tilde{q}^j}{\partial q^1} \mathrm{d}\tilde{p}_j\right) \wedge ... \wedge \left(\sum_{j=1}^{n} \frac{\partial \tilde{q}^j}{\partial q^{n+1}} \mathrm{d}\tilde{p}_j\right),
$$

is still a wedge product of n sums. So by introducing a notation

$$
\det\left(\left.\frac{\partial \tilde{q}^j}{\partial q^i}\right|_k^r\right)
$$

for a determinant of $(n \times n)$ submatrix of $(n+1 \times n+1)$ matrix from which the kth column and rth row have been omitted, makes the induction assumption applicable. In this submatrix notation (5.10) becomes

$$
\det\left(\left.\frac{\partial \tilde{q}^j}{\partial q^i}\right|_{n+1}^{n+1}\right) \mathrm{d}\tilde{p}_1 \wedge ... \wedge \mathrm{d}\tilde{p}_n \wedge (-1)^{(n+1)-(n+1)} \frac{\partial \tilde{q}^{n+1}}{\partial q^{n+1}} \mathrm{d}\tilde{p}_{n+1}
$$

and more generally it holds that

$$\left(\sum_{j=1}^{n}\frac{\partial\tilde{q}^j}{\partial q^1}\mathrm{d}\tilde{p}_j\right)\wedge\ldots\wedge\left(\sum_{j=1}^{n}\frac{\partial\tilde{q}^j}{\partial q^{n+1}}\mathrm{d}\tilde{p}_j\right)\wedge(-1)^{(n+1)-i}\frac{\partial\tilde{q}^{n+1}}{\partial q^i}\mathrm{d}\tilde{p}_{n+1}$$
$$=\det\left(\left.\frac{\partial\tilde{q}^j}{\partial q^i}\right|_i^{n+1}\right)\mathrm{d}\tilde{p}_1\wedge\ldots\wedge\mathrm{d}\tilde{p}_n\wedge(-1)^{(n+1)-i}\frac{\partial\tilde{q}^{n+1}}{\partial q^i}\mathrm{d}\tilde{p}_{n+1}.$$

Then using the submatrix notation to the whole sum of $n+1$ forms gives expression:

$$\begin{aligned}
&\det\left(\left.\frac{\partial\tilde{q}^j}{\partial q^i}\right|_1^{n+1}\right)\mathrm{d}\tilde{p}_1\wedge\ldots\wedge\mathrm{d}\tilde{p}_n\wedge(-1)^{(n+1)-1}\frac{\partial\tilde{q}^{n+1}}{\partial q^1}\mathrm{d}\tilde{p}_{n+1}\\
&+\det\left(\left.\frac{\partial\tilde{q}^j}{\partial q^i}\right|_2^{n+1}\right)\mathrm{d}\tilde{p}_1\wedge\ldots\wedge\mathrm{d}\tilde{p}_n\wedge(-1)^{(n+1)-2}\frac{\partial\tilde{q}^{n+1}}{\partial q^2}\mathrm{d}\tilde{p}_{n+1}\\
&+\ldots+\\
&+\det\left(\left.\frac{\partial\tilde{q}^j}{\partial q^i}\right|_i^{n+1}\right)\mathrm{d}\tilde{p}_1\wedge\ldots\wedge\mathrm{d}\tilde{p}_n\wedge(-1)^{(n+1)-i}\frac{\partial\tilde{q}^{n+1}}{\partial q^i}\mathrm{d}\tilde{p}_{n+1}\\
&+\ldots+\\
&+\det\left(\left.\frac{\partial\tilde{q}^j}{\partial q^i}\right|_n^{n+1}\right)\mathrm{d}\tilde{p}_1\wedge\ldots\wedge\mathrm{d}\tilde{p}_n\wedge(-1)^{(n+1)-n}\frac{\partial\tilde{q}^{n+1}}{\partial q^n}\mathrm{d}\tilde{p}_{n+1}\\
&+\det\left(\left.\frac{\partial\tilde{q}^j}{\partial q^i}\right|_{n+1}^{n+1}\right)\mathrm{d}\tilde{p}_1\wedge\ldots\wedge\mathrm{d}\tilde{p}_n\wedge(-1)^{(n+1)-(n+1)}\frac{\partial\tilde{q}^{n+1}}{\partial q^{n+1}}\mathrm{d}\tilde{p}_{n+1},
\end{aligned}$$

which after moving the coefficients of the covectors $\mathrm{d}\tilde{p}_{n+1}$ to the front, gives the following sum

$$
\begin{aligned}
&(-1)^{(n+1)-1}\frac{\partial\tilde{q}^{n+1}}{\partial q^1}\cdot\det\left(\frac{\partial\tilde{q}^j}{\partial q^i}\bigg|_1^{n+1}\right)\mathrm{d}\tilde{p}_1\wedge\ldots\wedge\mathrm{d}\tilde{p}_n\wedge\mathrm{d}\tilde{p}_{n+1}\\
&+(-1)^{(n+1)-2}\frac{\partial\tilde{q}^{n+1}}{\partial q^2}\cdot\det\left(\frac{\partial\tilde{q}^j}{\partial q^i}\bigg|_2^{n+1}\right)\mathrm{d}\tilde{p}_1\wedge\ldots\wedge\mathrm{d}\tilde{p}_n\wedge\mathrm{d}\tilde{p}_{n+1}\\
&+\ldots+\\
&+(-1)^{(n+1)-i}\frac{\partial\tilde{q}^{n+1}}{\partial q^i}\cdot\det\left(\frac{\partial\tilde{q}^j}{\partial q^i}\bigg|_i^{n+1}\right)\mathrm{d}\tilde{p}_1\wedge\ldots\wedge\mathrm{d}\tilde{p}_n\wedge\mathrm{d}\tilde{p}_{n+1}\\
&+\ldots+\\
&+(-1)^{(n+1)-n}\frac{\partial\tilde{q}^{n+1}}{\partial q^n}\cdot\det\left(\frac{\partial\tilde{q}^j}{\partial q^i}\bigg|_n^{n+1}\right)\mathrm{d}\tilde{p}_1\wedge\ldots\wedge\mathrm{d}\tilde{p}_n\wedge\mathrm{d}\tilde{p}_{n+1}\\
&+(-1)^{(n+1)-(n+1)}\frac{\partial\tilde{q}^{n+1}}{\partial q^{n+1}}\cdot\det\left(\frac{\partial\tilde{q}^j}{\partial q^i}\bigg|_{n+1}^{n+1}\right)\mathrm{d}\tilde{p}_1\wedge\ldots\wedge\mathrm{d}\tilde{p}_n\wedge\mathrm{d}\tilde{p}_{n+1},
\end{aligned}
$$

where the coefficients marked orange are exactly the Laplace expansion of an $(n+1\times n+1)$ matrix, and therefore

$$
\mathrm{d}p_1\wedge\ldots\wedge\mathrm{d}p_n\wedge\mathrm{d}p_{n+1}=\det\left(\frac{\partial\tilde{q}^j}{\partial q^i}\right)\mathrm{d}\tilde{p}_1\wedge\ldots\wedge\mathrm{d}\tilde{p}_{n+1}.
$$

□

Bibliography

[1] John M.Lee, *Introduction to smooth manifolds*, 2013, 2nd Edition, Graduate Texts in Mathematics 218, Springer-Verlag, New York.

[2] Betancourt, Michael; Byrne, Simon; Livingstone, Sam; Girolami, Mark. *The geometric foundations of Hamiltonian Monte Carlo.* Bernoulli 23 (2017), no. 4A, 2257–2298.

[3] Gamerman, D. and Lopes, H.F., *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*, 2006, 2nd Edition, Texts in Statistical Science, CRC Press.

[4] Michael Betancourt. *Conceptual introduction to Hamiltonian Monte Carlo.* 2017, arXiv:1701.02434v2.

[5] Ricardo Mañé, *Ergodic Theory and Differentiable Dynamics*, 1987, A series of modern surveys of mathematics, Springer-Verlag Berlin Heidelberg.

[6] Rudolf, Daniel. *An introduction to Markov chain Monte Carlo on general state spaces*, lecture notes, 2012, University of New South Wales.

[7] Diaconis, P. and Freedman, D., 1999. *Iterated random functions.* SIAM review, 41(1), pp.45-76.

[8] Durmus, A., Moulines, E. and Saksman, E., 2017. *On the convergence of Hamiltonian Honte Carlo.* ArXiv:1705.00166.

[9] Bou-Rabee, N., Eberle, A. and Zimmer, R., 2018. *Coupling and convergence for Hamiltonian monte carlo*, arXiv:1805.00452.

[10] Ernst Hairer, *Lecture 2: Symplectic integrators*, Geometric Numerical Integration, TU Munich 2010

[11] Hoffman, M.D. and Gelman, A. *The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo.* 2014, J. Mach. Learn. Res., 15(1), pp.1593-1623.

[12] Livingstone, S., Betancourt, M., Byrne, S. and Girolami, M. *On the geometric ergodicity of Hamiltonian Monte Carlo.* 2019, Bernoulli, 25(4A), pp.3109-3138.

[13] Nishimura, A. and Dunson, D. *Geometrically tempered Hamiltonian Honte Carlo.* 2016, arXiv:1604.00872.

[14] Graham, M.M. and Storkey, A.J. *Continuously tempered Hamiltonian Monte Carlo.* UAI 2017.

[15] Luo, R., Wang, J., Yang, Y., Zhu, Z. and Wang, J. *Thermostat-assisted continuously-tempered Hamiltonian Monte Carlo for Bayesian learning.* NeurIPS 2018.

[16] Beskos, A., Pinski, F.J., Sanz-Serna, J.M. and Stuart, A.M. *Hybrid Monte Carlo on Hilbert spaces.* Stochastic Processes and their Applications. 2011, vol. 121, issue 10, pp. 2201-2230.