Mathematical
Surveys
and
Monographs

Volume 74

An Introduction to the Analysis of Paths on a Riemannian Manifold

Daniel W. Stroock

American Mathematical Society

Selected Titles in This Series

74 **Daniel W. Stroock**, An introduction to the analysis of paths on a Riemannian manifold, 2000

73 **John Locker**, Spectral theory of non-self-adjoint two-point differential operators, 2000

72 **Gerald Teschl**, Jacobi operators and completely integrable nonlinear lattices, 1999

71 **Lajos Pukánszky**, Characters of connected Lie groups, 1999

70 **Carmen Chicone and Yuri Latushkin**, Evolution semigroups in dynamical systems and differential equations, 1999

69 **C. T. C. Wall (A. A. Ranicki, Editor)**, Surgery on compact manifolds, second edition, 1999

68 **David A. Cox and Sheldon Katz**, Mirror symmetry and algebraic geometry, 1999

67 **A. Borel and N. Wallach**, Continuous cohomology, discrete subgroups, and representations of reductive groups, second edition, 2000

66 **Yu. Ilyashenko and Weigu Li**, Nonlocal bifurcations, 1999

65 **Carl Faith**, Rings and things and a fine array of twentieth century associative algebra, 1999

64 **Rene A. Carmona and Boris Rozovskii, Editors**, Stochastic partial differential equations: Six perspectives, 1999

63 **Mark Hovey**, Model categories, 1999

62 **Vladimir I. Bogachev**, Gaussian measures, 1998

61 **W. Norrie Everitt and Lawrence Markus**, Boundary value problems and symplectic algebra for ordinary differential and quasi-differential operators, 1999

60 **Iain Raeburn and Dana P. Williams**, Morita equivalence and continuous-trace C^*-algebras, 1998

59 **Paul Howard and Jean E. Rubin**, Consequences of the axiom of choice, 1998

58 **Pavel I. Etingof, Igor B. Frenkel, and Alexander A. Kirillov, Jr.**, Lectures on representation theory and Knizhnik-Zamolodchikov equations, 1998

57 **Marc Levine**, Mixed motives, 1998

56 **Leonid I. Korogodski and Yan S. Soibelman**, Algebras of functions on quantum groups: Part I, 1998

55 **J. Scott Carter and Masahico Saito**, Knotted surfaces and their diagrams, 1998

54 **Casper Goffman, Togo Nishiura, and Daniel Waterman**, Homeomorphisms in analysis, 1997

53 **Andreas Kriegl and Peter W. Michor**, The convenient setting of global analysis, 1997

52 **V. A. Kozlov, V. G. Maz′ya, and J. Rossmann**, Elliptic boundary value problems in domains with point singularities, 1997

51 **Jan Malý and William P. Ziemer**, Fine regularity of solutions of elliptic partial differential equations, 1997

50 **Jon Aaronson**, An introduction to infinite ergodic theory, 1997

49 **R. E. Showalter**, Monotone operators in Banach space and nonlinear partial differential equations, 1997

48 **Paul-Jean Cahen and Jean-Luc Chabert**, Integer-valued polynomials, 1997

47 **A. D. Elmendorf, I. Kriz, M. A. Mandell, and J. P. May (with an appendix by M. Cole)**, Rings, modules, and algebras in stable homotopy theory, 1997

46 **Stephen Lipscomb**, Symmetric inverse semigroups, 1996

45 **George M. Bergman and Adam O. Hausknecht**, Cogroups and co-rings in categories of associative rings, 1996

For a complete list of titles in this series, visit the
AMS Bookstore at **www.ams.org/bookstore/**.

An Introduction to the Analysis of Paths on a Riemannian Manifold

Mathematical
Surveys
and
Monographs

Volume 74

An Introduction to the Analysis of Paths on a Riemannian Manifold

Daniel W. Stroock

American Mathematical Society

1991 *Mathematics Subject Classification.* Primary 60J65; Secondary 60J60, 60D05.

ABSTRACT. This book provides an introduction to Brownian motion on a Riemannian manifold. Although the reader is expected to have some familiarity with both probability theory and differential geometry, the author has attempted to make the book self-contained. Thus, whenever technically demanding topics are introduced (e.g., Brownian motion in probability theory and the orthonormal bundle of frames in differential geometry), he has provided some background. His hope is that the book will be accessible to anyone who has had semester courses in probability theory and differential geometry at the graduate level.

Library of Congress Cataloging-in-Publication Data

Stroock, Daniel W.
An introduction to the analysis of paths on a Riemannian manifold / Daniel W. Stroock.
p. cm. — (Mathematical surveys and monographs, ISSN 0076-5376 ; v. 74)
Includes bibliographical references and index.
ISBN 0-8218-2020-6 (alk. paper)
1. Riemannian manifolds. 2. Brownian motion processes. I. Title. II. Mathematical surveys and monographs ; no. 74.
QA649 .S76 1999
516.3′73–dc21 99-044329

∞ The paper used in this book is acid-free and falls within the guidelines
established to ensure permanence and durability.
Visit the AMS home page at URL: http://www.ams.org/

10 9 8 7 6 5 4 3 2 1 05 04 03 02 01 00

This book is dedicated to my friends and mentors:

Paul G. Malliavin and *Shing Tung Yau*

Contents

Preface

It was at a C.I.M.E. conference at the Palazzo in Cortona during the summer of 1978 that my eyes were opened to Malliavin's multi-tiered mansion in which Brownian motion on a Riemannian manifold resides. There, in the Palazzo's beautiful ballroom with its tiny blackboard presided over by Cleopatra and her adder, Malliavin held his audience in thrall with tales whose comprehension demanded simultaneous appreciation of the "upstairs story," the "downstairs story," and the profound influence that events on either exercise on the other. I have to admit that I could not have said with certainty on exactly which "level" a given event was transpiring. Indeed, at first I thought that there were only two "levels:" the upper one where Wiener measure lives and the lower one which is the manifold where the Brownian motion is taking place.

My confusion about this critical point was a direct consequence of my nearly perfect ignorance of differential geometry. In particular, because I had no idea what it was, Malliavin's frequent references to an intermediate level called the "bundle of orthonormal frames" were lost on me. Such matters are not broached in the first ten pages of even the most ambitious introductory texts about Riemannian geometry, and the first ten pages is as far as I had ever penetrated into the many differential geometry books which I had failed to read. Nor were Malliavin's intriguing lectures sufficient to persuade me to mend my ways immediately. Indeed, another fifteen years passed before my joint work with first Shigeo Kusuoka and then Ognian Enchev finally convinced me that the pain resulting from not learning more differential geometry would inevitably exceed the pain of mastering more than the first ten pages of at least one differential geometry text. Thus, about five years ago I forced myself to come to terms with Bishop and Crittenden's remarkably concise text [2]. My choice was dictated by two considerations: first, my collaborator Enchev had already assimilated the material in this book and I did not want to fall too far behind; secondly, Bishop and Crittenden emphasize the role of the bundle of orthonormal frames, and Malliavin had already alerted me to the advantages of this perspective. Of course, once I had taken the plunge, I delved into several other sources. In fact, the citations in this text give a reasonably accurate map of where I learned what.

Having benefitted from the efforts of differential geometers to explain their subject to me through their writings, I decided to reciprocate by writing this book, which is my attempt to explain my subject to them. With this in mind, I have tried to minimize the weight of "probabilistic" baggage which my readers must bring to a reading of this book. Further, wherever the option was available, I have chosen to emphasize the geometric over the stochastic aspect of the topic at hand. In particular, I never have made explicit use here of Itô's stochastic calculus. In spite of the grand and beautiful edifice erected by L. Schwartz, R. Darling, P.A. Meyer, and others (cf. [14] for an excellent explanation of their ideas or Ikeda and Watanabe's famous [22] for a more standard treatment) to convince me and the world otherwise, I remain firmly convinced that Stratonovich calculus is the calculus of choice if one wants to maximize ones geometric insight into stochastic analysis on differentiable manifolds. Thus, I have, from the outset, solved all "my stochastic integral equations" (the quotation marks are here because this is the last time that the term "stochastic integral equation" makes an appearance in this book) by passing to limits after mollification. My hope was that this procedure will make the book more accessible to readers who have not been reared in the probabilistic tradition. My fear is that I may very well have produced a book which is incomprehensible equally to the probabilistic and differential geometric communities. Be that as it may, here is a summary of the material which I have tried to convey.

Because I did not want to assume that my reader is acquainted with Wiener measure, I have devoted Chapter 1 to the construction of Wiener measure and a brief resume of some of its properties. There are, by now, a myriad construction methods. The one which I have chosen is basically the one given by P. Lévy. Not only is Lévy's construction stunningly beautiful, it has the advantage that, in some sense, it sets the pattern for all the other constructions which follow.

Following Itô's ideas, but not his procedure, I use the techniques, orignally explained in [39], to show in Chapters 2 and 3 how one can massage Wiener paths into the paths of more general diffusions on $\mathbb{R}^N$. Chapter 2 covers the basic case, the one in which everything is sufficiently bounded that no problems about possible explosion ever arise. In Chapter 3, it is shown that much of what is done in Chapter 2 continues to hold even after the boundedness assumptions are removed. In addition, Chapter 3 addresses several other topics of importance, chief of which are subordination and invariant measures.

Differentiable manifolds make their initial appearance in Chapter 4, where they appear as an embedded submanifold M of $\mathbb{R}^N$. First it is shown that quite general diffusions on M can be viewed as special cases of the diffusions constructed in Chapters 2 and 3. Second, when M is given the Riemannian structure which it inherits from $\mathbb{R}^N$, it is shown that the Brownian motion on M can be realized by "projecting" Wiener paths from the ambient $\mathbb{R}^N$ onto M.[1] The unabashedly extrinsic ideas initiated in Chapter 4 are developed further in Chapters 5 and 6. Specifically, curvature considerations are introduced in Chapter 5, where, in connection with Yau's non-explosion criterion,

[1] So far as I know, the first time that such a construction of Brownian motion appears is when, as Itô pointed out, I had stumbled upon it in [38] for the 2-sphere in $\mathbb{R}^3$. Subsequently, John Lewis [26] realized that the same construction works in general, although he lost the interpretation in terms of a projection. Nonetheless, the projection reappeared in the treatment given by Chris Rogers and David Williams [33].

I present the first evidence that Ricci curvature has a lot to say about the behavior of Brownian paths. Further evidence of the same fact is provided in Chapter 6, where I prove Bochner's identity in an integrated form which leads to a beautiful interpretation given by J.-M. Bismut in [3].

The rest of the book takes an intrinsic point of view. In Chapter 7, it is explained how the material in Chapters 2, 3, and 4 transfers, without difficulty, to the setting of an abstract differentiable manifold M. In particular, Chapter 7 ends with a "dirty," hands-on construction of Brownian motion via localization. In order to prepare the way for the intrinsic construction of Brownian motion due to Eells, Elworthy, and Malliavin (cf. [11] and [29]), Chapter 8 starts with a quick summary of the basic facts about the bundle $\mathcal{O}(M)$ of orthonormal frames, gives the E-E-M construction of Brownian motion as the projection from $\mathcal{O}(M)$ to M of the diffusion on $\mathcal{O}(M)$ associated with Bochner's Laplacian, and ends with a demonstration that all the essentially intrinsic results proved earlier about Brownian motion on a submanifold are, if anything, easier to understand in this abstract setting.

Chapter 9 is something of a digression. The idea is to expose how systematic use of normal coordinates enters into the study of Brownian motion on a manifold. Not surprisingly, the applications are strictly local. For example, it is shown how familiar expansions of the metric in normal coordinates are manifested in the computation of the exit time and exit place of Brownian motion from very small balls.

Finally, in the concluding chapter I take up the topic which originally stimulated my own interest in Brownian paths on Riemannian manifolds. Namely, for many years I worked on a set of ideas which I dubbed the *Malliavin calculus*. The essential, unifying theme of these ideas is that useful analytic information can be obtained from doing differential calculus in pathspace. More precisely, by perturbing paths and examining the infinitesimal response of their distribution to the perturbation, one can gain insight into various analytic quantities which are representable in terms of distribution of those paths. I had (most successfully with Shigeo Kusuoka) practiced this art in the Euclidean context. Around the same time, Jean–Michel Bismut (cf. [3]) was taking the initial steps which are necessary if one wants to do the same thing in a differential geometric setting. Somewhat later, Bismut's program was given an enormous boost by Bruce Driver's key article [9]. Motivated, at least in part, by the desire not to read all 104 pages of Driver's paper, Ognian Enchev and I embarked on a program to obtain Driver's conclusions on our own, and Chapter 10 is derived from the paper [15] which grew out of our efforts.

Finally, I have to recognize the critical role that my friend S.–T. Yau has played in all this. In particular, Yau consistently challenged me to come up with something that probability theory could do that Yau himself could not. Of course, I knew all along that such an example does not exist, but I was damned if I would tell Yau. Now I have.

Daniel W. Stroock, June 1999

CHAPTER 1

Brownian Motion in Euclidean Space

1.1 Wiener Measure

In order for a mathematician to take A. Einstein's 1905 article [12] seriously, he should feel obliged to begin by doing what N. Wiener did in his 1923 article [45]. Namely, he should convince himself that there is one, and only one, Borel, probability measure $\mu_{\mathbb{R}}$ on the Frechét space[1]

$$\mathfrak{W}(\mathbb{R}) \equiv \{w \in C([0,\infty);\mathbb{R}) : w(0) = 0\}$$

with the properties that, for each $t \in [0,\infty)$ and $\tau > 0$, the *increment* $w \in \mathfrak{W}(\mathbb{R}) \longmapsto w(t+\tau) - w(t) \in \mathbb{R}$ under $\mu_{\mathbb{R}}$ is a *centered Gaussian random variable* with variance τ which is independent of the σ-algebra $\mathcal{B}_t \equiv \sigma(\{w(s) : s \in [0,t]\})$.[2] Equivalently, for each $n \geq 1$, $0 = t_0 < t_1 < \cdots < t_n$, $\Gamma_1, \ldots, \Gamma_n \in \mathcal{B}_{\mathbb{R}}$:[3]

$$\begin{aligned} &\mu_{\mathbb{R}}(\{w : w(t_1) - w(t_0) \in \Gamma_1, \ldots, w(t_n) - w(t_{n-1}) \in \Gamma_n\}) \\ &\qquad = \prod_{m=1}^{n} \int_{\Gamma_m} g_{\tau_m}(\eta_m)\, d\eta_m, \end{aligned} \tag{1.1}$$

where

$$g_\tau(\eta) \equiv \frac{1}{\sqrt{2\pi\tau}} \exp\left(-\frac{y^2}{2\tau}\right), \quad (\tau, y) \in (0,\infty) \times \mathbb{R} \tag{1.2}$$

and $\tau_m = t_m - t_{m-1}$. Actually, even though it requires some effort to produce one such $\mu_{\mathbb{R}}$, it is essentially trivial to check that there is at most one. Indeed, (1.1) obviously determines $\mu_{\mathbb{R}}$ on $\sigma(\{w(s) : s \geq 0\})$, and so the only thing that has to be checked (cf. §3.3 in [35]) is that $\mathcal{B}_{\mathfrak{W}(\mathbb{R})} = \sigma(\{w(s) : s \geq 0\})$.

The reason why the mathematically inclined reader of Einstein should want $\mu_{\mathbb{R}}$ to exist is that, according to Einstein, $\mu_{\mathbb{R}}$-typical paths are the trajectories

[1] Here we are thinking of the Frechét metric which corresponds to uniform convergence on compact intervals.

[2] When $\mathcal{F}$ is a family of functions on some space Ω, we will use $\sigma(\mathcal{F})$ to denote the smallest σ-algebra over Ω with respect to which every element of $\mathcal{F}$ is measurable.

[3] Given a separable metric space E, we use $\mathcal{B}_E$ to denote the Borel field over E.

of *Brownian particles.*[4] Moreover, even if one ignores its physical implications, the existence of $\mu_{\mathbb{R}}$ raises a question of fundamental mathematical interest: can one put non-trivial, countably additive measures on an infinite dimensional spaces?

1.1.1. Deconstructing Brownian Paths. The construction of Wiener measure which we will give in the next subsection is not any one of the three suggested by Wiener in [45]. Instead, it is derived from the ideas of P. Lévy and is based on the intuition explained here.

Assume that $\mu_{\mathbb{R}}$ exists. It is then clear that there are lots and lots of mutually independent, centered, Gaussian random variables floating around. Indeed, for any $n \geq 1$ and $0 \leq t_0 < \cdots < t_n$, the increments $\{w(t_m) - w(t_{m-1}) : 1 \leq m \leq n\}$ will be mutually independent, centered, Gaussian random variables. Since one understands how to deal with and construct mutually independent random variables, it is reasonable to seek a clever way to choose a countable set of increments from which the whole path w can be re-constructed via an elementary, deterministic procedure. The point is that, once such a scheme has been found, it should be possible to construct $\mu_{\mathbb{R}}$. Namely, one will use standard measure theoretic methods to construct an ample supply of mutually independent, centered Gaussian random variables, which one will then plug into the "re-construction" procedure.

Everything which follows relies heavily on the peculiar properties possessed by linear families of centered, Gaussian random variable. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we will say that $\mathfrak{G} \subseteq L^2(\mathbb{P}; \mathbb{R})$ is a (centered) *Gaussian family* if it is a linear subspace each of whose elements is a centered, Gaussian random variable. The essential facts for us are summarized in the following lemma.

1.3 LEMMA. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathfrak{G}$ be a Gaussian family in $L^2(\mathbb{P}; \mathbb{R})$. Then the closure $\bar{\mathfrak{G}}$ of $\mathfrak{G}$ in $L^2(\mathbb{P}; \mathbb{R})$ is also a Gaussian family. Moreover, if K is a non-empty subset of $\mathfrak{G}$, then $\sigma(K)$ is independent of $\sigma(K^\perp \cap \mathfrak{G})$. Finally, if $\mathfrak{H}$ is a second Gaussian family in $L^2(\mathbb{P}; \mathbb{R})$ and if $\sigma(\mathfrak{G})$ is independent of $\sigma(\mathfrak{H})$, then $\mathfrak{G} + \mathfrak{H}$ is again a Gaussian family.*

PROOF: All of these statements are applications of the fact that $Y \in L^2(\mathbb{P}; \mathbb{R})$ is a centered Gaussian random variable if and only if[5]

$$\mathbb{E}^{\mathbb{P}}\left[e^{\sqrt{-1}\,\xi Y}\right] = \exp\left(-\frac{\xi^2}{2}\mathbb{E}^{\mathbb{P}}\left[Y^2\right]\right), \quad \xi \in \mathbb{R}.$$

[4] This term is used rather loosely to describe particles whose motion displays the sort of chaotic behavior observed by the nineteenth century botanist R. Brown, who was looking at pollen under a microscope. Einstein and (more or less simultaneously) M. Smoluchowski accounted for Brown's observations as a manifestation of the kinetic theory of gases. Thus, unless $\mu_{\mathbb{R}}$ exists, Einstein and Smoluchowski's whole picture wonld fall under the shadow of considerable mathematical doubt, a circumstance which might not have particnlarly disturbed Einstein but would probably have caused Smoluchowski severe distress.

[5] Below, and throughout, I have adopted the usual probabilistic convention of identifying integrals as *expectation values* and therefore using $\mathbb{E}^{\mathbb{P}}[F]$ to denote $\int F\,d\mathbb{P} = \int_\Omega F(\omega)\,\mathbb{P}(d\omega)$ when doing integration with respect to a generic probability measure $\mathbb{P}$.

Thus, if $\{Y_n\}_1^\infty \subseteq \mathfrak{G}$ converges in $L^2(\mathbb{P};\mathbb{R})$ to Y, then

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\left[e^{\sqrt{-1}\,\xi Y}\right] &= \lim_{n\to\infty}\mathbb{E}^{\mathbb{P}}\left[e^{\sqrt{-1}\,\xi Y_n}\right] \\ &= \lim_{n\to\infty}\exp\left(-\frac{\xi^2}{2}\mathbb{E}^{\mathbb{P}}\left[Y_n^2\right]\right) = \exp\left(-\frac{\xi^2}{2}\mathbb{E}^{\mathbb{P}}\left[Y^2\right]\right),\end{aligned}$$

which means that Y is also a centered, Gaussian random variable. Similarly, to prove the third statement, all that one needs to show is that, for $X \in \mathfrak{G}$ and $Y \in \mathfrak{H}$, $X + Y$ is a centered, Gaussian random variable. But, because X and Y are independent and centered,

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\left[e^{\sqrt{-1}\,\xi(X+Y)}\right] &= \mathbb{E}^{\mathbb{P}}\left[e^{\sqrt{-1}\,\xi X}\right]\mathbb{E}^{\mathbb{P}}\left[e^{\sqrt{-1}\,\xi Y}\right] \\ &= \exp\left(-\frac{\xi^2}{2}\left(\mathbb{E}^{\mathbb{P}}\left[X^2\right]+\mathbb{E}^{\mathbb{P}}\left[Y^2\right]\right)\right) = \exp\left(-\frac{\xi^2}{2}\mathbb{E}^{\mathbb{P}}\left[(X+Y)^2\right]\right).\end{aligned}$$

Finally, to prove the second statement, it suffices[6] to show that for any finite collections $\{X_1,\dots,X_m\} \subseteq K$ and $Y_1,\dots,Y_n \subseteq K^\perp \cap \mathfrak{G}$,

$$\begin{aligned}&\mathbb{E}^{\mathbb{P}}\left[\exp\left(\sqrt{-1}\sum_{i=1}^m \xi_i X_i\right)\exp\left(\sqrt{-1}\sum_{j=1}^n \eta_j Y_j\right)\right] \\ &\qquad = \mathbb{E}^{\mathbb{P}}\left[\exp\left(\sqrt{-1}\sum_{i=1}^m \xi_i X_i\right)\right]\mathbb{E}^{\mathbb{P}}\left[\exp\left(\sqrt{-1}\sum_{j=1}^n \eta_j Y_j\right)\right]\end{aligned}$$

for all $(\xi_1,\dots,\xi_m) \in \mathbb{R}^m$ and $(\eta_1,\dots,\eta_n) \in \mathbb{R}^n$. In the present situation, this is tantamount to checking that

$$\begin{aligned}&\mathbb{E}^{\mathbb{P}}\left[\exp\left(\sqrt{-1}\sum_{i=1}^m \xi_i X_i\right)\exp\left(\sqrt{-1}\sum_{j=1}^n \eta_j Y_j\right)\right] \\ &\qquad = \exp\left(-\frac{1}{2}\left(\mathbb{E}^{\mathbb{P}}\left[\left(\sum_{i=1}^m \xi_i X_i\right)^2\right] + \mathbb{E}^{\mathbb{P}}\left[\left(\sum_{j=1}^n \eta_j Y_j\right)^2\right]\right)\right),\end{aligned}$$

which, in turn, follows from the fact that $\sum_{i=1}^m \xi_i X_i$ is orthogonal to $\sum_{j=1}^n \eta_j Y_j$. □

The immediate relevance to us of Lemma 1.3 is that the increments

$$\{w(t) - w(s) : 0 \le s < t\}$$

[6] The reader who wants a quick review of the basic facts about independent random variables might want to take a look at §1.1 in [35]. In particnlar, Exercise 1.1.13 there is relevant here.

form a Gaussian family in $L^2(\mu_{\mathbb{R}};\mathbb{R})$. Armed with this observation, we return to the problem at hand. As a first guess, one might try looking at the increments

$$\Delta_{m,0}w \equiv w(m+1) - w(m), \quad m \in \mathbb{N},$$

and using linear interpolation to construct a path w_0. That is, take $w_0(t) = w(m) + (t-m)\Delta_{m,0}w$ for $m \geq 0$ and $t \in [m, m+1]$. Of course, $w = w_0$ at integer times. However, if t is not an integer, then $\mu_{\mathbb{R}}\big(w(t) = w_0(t)\big) = 0$. For example, because $w_0\big((2\ell+1)2^{-1}\big) = \frac{1}{2}\big(w(\ell+1) - w(\ell)\big)$,

$$\begin{aligned} D_{\ell,1}w &\equiv w\big((2\ell+1)2^{-1}\big) - w_0\big((2\ell+1)2^{-1}\big) \\ &= \frac{\Delta_{2\ell,1}w - \Delta_{2\ell+1,1}w}{2}, \end{aligned}$$

where

$$\Delta_{m,1}w \equiv w\big((m+1)2^{-1}\big) - w\big(m2^{-1}\big).$$

Since the $\Delta_{m,1}w$'s are mutually independent, centered Gaussian random variables with variance $\frac{1}{2}$, $D_{\ell,1}w$ is a centered Gaussian random variable with variance $\frac{1}{4}$. Moreover, each $D_{\ell,1}w$ is orthogonal to all of the $\Delta_{m,0}w$'s, and therefore, by the second part of Lemma 1.3, the σ-algebras

$$\sigma\Big(\big\{D_{\ell,1}w : \ell \in \mathbb{N}\big\}\Big) \quad \text{and} \quad \sigma\Big(\big\{\Delta_{m,0}w : m \in \mathbb{N}\big\}\Big)$$

are independent. Thus, there simply is not enough *randomness* in the $\Delta_{m,0}w$'s, and we will have to throw in the $D_{\ell,1}w$'s if we are going to have any chance of reconstructing w.

With the preceding in mind, for each $n \in \mathbb{N}$, we introduce the increments

$$\Delta_{m,n}w \equiv w\big((m+1)2^{-n}\big) - w\big(m2^{-n}\big), \quad m \in \mathbb{N},$$

construct the polygonal path w_n so that

$$w_n(t) = w\big(m2^{-n}\big) + \big(2^n t - m\big)\Delta_{m,n}w \text{ for } m \in \mathbb{N} \;\&\; t \in \big[m2^{-n}, (m+1)2^{-n}\big],$$

take

$$D_{\ell,n}w \equiv w\big((2\ell+1)2^{-n}\big) - w_n\big((2\ell+1)2^{-n}\big) = \frac{\Delta_{2\ell,n}w - \Delta_{2\ell+1,n}w}{2},$$

when $n \geq 1$, and define the random variables

$$(1.4) \qquad Y_{m2^{-n}}w = \begin{cases} \Delta_{m,0}w & \text{if } n = 0 \\ 2^{\frac{n+1}{2}} D_{\ell,n}w & \text{if } n \geq 1 \text{ and } m = 2\ell+1. \end{cases}$$

Using Lemma 1.3 as we did above, one finds that, under $\mu_{\mathbb{R}}$,

$$\big\{Y_{m2^{-n}}w : (m,n) \in \mathbb{N}^2\big\}$$

constitutes a family of mutually independent, centered Gaussian random variables with variance 1. Moreover, one can easily reconstruct w from the $Y_{m2^{-n}}$'s. Namely, $w_0(0) = 0$

$$
\begin{aligned}
&w_0(t) = w_0(m) + (t-m)Y_m w \quad \text{for } m \in \mathbb{N} \text{ and } t \in [m, m+1], \\
&w_n(t) = w_{n-1}(t) + 2^{-\frac{n+1}{2}}\big(1 - |2^n t - (2m+1)|\big)Y_{(2m+1)2^{-n}} w \\
&\qquad \text{for } n \geq 1,\ m \in \mathbb{N},\ \text{and } t \in \big[m2^{-n+1}, (m+1)2^{-n+1}\big],
\end{aligned} \tag{1.5}
$$

and $w_n \longrightarrow w$ uniformly on compacts.

1.1.2. Lévy's Construction. It should be now clear how to proceed. First, one constructs a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which can support a countably infinite family of mutually independent random variables.[7] One should then label these random as $Y_{m2^{-n}}$ for $(m, n) \in \mathbb{N}^2$. Next, use (1.5) as the template for constructing polygonal paths w_n:

$$
\begin{aligned}
w_0(0) = 0 \quad \text{and} \quad & w_0(t) = w_0(m) + (t-m)Y_m \quad \text{for } m \in \mathbb{N} \ \&\ t \in [m, m+1], \\
w_n(t) &= w_{n-1}(t) + 2^{-\frac{n+1}{2}}\big(1 - |2^n t - (2m+1)|\big)Y_{(2m+1)2^{-n}} \\
&\text{for } n \geq 1,\ m \in \mathbb{N},\ \text{and } t \in \big[m2^{-n+1}, (m+1)2^{-n+1}\big].
\end{aligned}
$$

Notice that $w_n\big((2m+1)2^{-n} \pm 2^{-n}\big) - w_n\big((2m+1)2^{-n}\big)$ equals

$$
-2^{-\frac{n+1}{2}} Y_{(2m+1)2^{-n}} \pm \frac{w_{n-1}\big((m+1)2^{-n+1}\big) - w_{n-1}\big(m2^{-n+1}\big)}{2}.
$$

In particular, with the help of Lemma 1.3, one can use induction on $n \in \mathbb{N}$ to check that, for each n,

$$
\big\{w_n\big((m+1)2^{-n}\big) - w_n\big(m2^{-n}\big) : m \in \mathbb{N}\big\}
$$

is a sequence of mutually independent, centered Gaussian random variables with variance 2^{-n}. Thus, if we can show that, as $n \to \infty$, the sequence $\{w_n\}_0^\infty$ is $\mathbb{P}$-almost surely convergent uniformly on compacts to a limit path w, then we can simply take $\mu_{\mathbb{R}}$ to be the distribution of a random variable of w under $\mathbb{P}$.[8] For this reason, consider $w_n - w_{n-1}$ and observe that

$$
\big|w_n(t) - w_{n-1}(t)\big| \leq 2^{-\frac{n+1}{2}} \big|Y_{(2m+1)2^{-n}}\big| \quad \text{for } t \in \big[m2^{-n}, (m+1)2^{-m}\big].
$$

[7] For example, one can (cf. Theorem 1.1.10 in [35]) take $\Omega = [0,1)$, $\mathcal{F} = \mathcal{B}_{[0,1)}$, and $\mathbb{P}$ to be the restriction of Lebesgue measure to $[0,1)$. Alternatively, one can use a product construction (cf. Exercise 1.1.14 in [35]).

[8] Given a random variable Ξ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in some measurable space $(E, \mathcal{B})$, probabilists call the *pushforward measure* $\Xi_*\mathbb{P} = \mathbb{P} \circ \Xi^{-1}$, given by $\mathbb{P} \circ \Xi^{-1}(\Gamma) = \mathbb{P}(\Xi \in \Gamma)$, the distribution of Ξ under $\mathbb{P}$. In the present setting, we are thinking of w as being a $\mathfrak{W}(\mathbb{R})$-valued random variable on Ω.

Hence, for any fixed $T \in (0,\infty)$,

$$\sup_{\tau\in[0,T]}\left|w_n(\tau) - w_{n-1}(\tau)\right| \le 2^{-\frac{n+1}{2}} \max_{m\le 2^{n-1}T}\left|Y_{(2m+1)2^{-n}}\right| \le 2^{-\frac{n+1}{2}}\left(\sum_{m\le 2^{n-1}T}\left|Y_{(2m+1)2^{-n}}\right|^4\right)^{\frac{1}{4}},$$

and so, by Jensen's inequality,

$$\mathbb{E}^{\mathbb{P}}\left[\sup_{\tau\in[0,T]}\left|w_n(\tau) - w_{n-1}(\tau)\right|\right] \le 2^{-\frac{n+1}{2}}\left(\sum_{m\le 2^{n-1}T}\mathbb{E}^{\mathbb{P}}\left[Y^4_{(2m+1)2^{-n}}\right]\right)^{\frac{1}{4}} \le \left(3(1+T)\right)^{\frac{1}{4}}2^{-\frac{n}{4}},$$

which is more than enough to show that

$$\sum_{n=1}^{\infty}\mathbb{E}^{\mathbb{P}}\left[\sup_{t\in[0,T]}\left|w_n(t) - w_{n-1}(t)\right|\right] < \infty.$$

Hence, standard measure theoretic machinery now applies and guarantees that, $\mathbb{P}$-almost surely, $\{w_n\}$ is Cauchy convergent in $\mathfrak{W}(\mathbb{R})$, and that is exactly what we wanted to show. In other words, if we use w to denote this limit, then as we said before, we can take $\mu_{\mathbb{R}}$ to be the distribution of w under $\mathbb{P}$.

1.1.3. Modulus of Continuity. By sharpening the preceding argument a little, one can estimate the modulus of continuity of Brownian paths. To see this, first observe for any $p \in [1,\infty)$ and $q \in (1,\infty)$ with $q \ge \frac{p}{2}$,[9]

$$\mathbb{E}^{\mathbb{P}}\left[\sup_{\tau\in[0,T]}\left|w_n(\tau) - w_{n-1}(\tau)\right|^p\right]^{\frac{1}{p}} \le C_q(1+T)^{\frac{1}{2q}}2^{-\frac{n}{2q'}},$$

where

$$C_q \equiv \left(\int_{\mathbb{R}}\xi^{2q}\, g_1(\xi)\, d\xi\right)^{\frac{1}{2q}}$$

and $q' = \frac{q}{q-1}$ is the Hölder conjugate of q. Indeed, the reasoning is the same as we used above when $p = 1$ and $q = 2$. Hence,

$$(*)\qquad \mathbb{E}^{\mathbb{P}}\left[\sup_{\tau\in[0,T]}\left|w(\tau) - w_n(\tau)\right|^p\right]^{\frac{1}{p}} \le C_q'(1+T)^{\frac{1}{2q}}2^{-\frac{n}{2q'}},$$

[9] In the hope of avoiding confusion which might arise from an attempt to parse the notation which follows, I point out that I often use $\mathbb{E}^{\mathbb{P}}[\cdots]^{\frac{1}{p}}$ when I mean $\left(\mathbb{E}^{\mathbb{P}}[\cdots]\right)^{\frac{1}{p}}$.

where $C'_q = C_q\left(1-2^{-\frac{1}{2q'}}\right)^{-1}$.

Next, let $0 \le s < t \le T$ be given, assume that $2^{-n-2} \le t-s \le 2^{-n-1}$ for some $n \in \mathbb{N}$, and choose $m \in \mathbb{N}$ so that $m \le 2^n s < 2^n t \le m+1$. Then (cf. the notation in §1.1.1, and remember that $w(m2^{-n}) = w_n(m2^{-n})$ for all $m \in \mathbb{N}$),

$$\begin{aligned}\left|w(t)-w(s)\right| &\le \left|w_n(t)-w_n(s)\right| + 2\sup_{\tau\in[0,T]}\left|w(\tau)-w_n(\tau)\right| \\ &= 2^n(t-s)\left|\Delta_{m,n}w\right| + 2\sup_{\tau\in[0,T]}\left|w(\tau)-w_n(\tau)\right| \\ &\le \frac{1}{2}\left|\Delta_{m,n}w\right| + 2\sup_{\tau\in[0,T]}\left|w(\tau)-w_n(\tau)\right|,\end{aligned}$$

and so, for any $\alpha > 0$,

$$\begin{aligned}&\sup_{\substack{0\le s<t\le T\\ t-s\le 1}} \frac{\left|w(t)-w(s)\right|}{(t-s)^\alpha} \\ &\qquad \le 4^\alpha \sum_{n=0}^\infty 2^{n\alpha}\left[\frac{1}{2}\max_{m\le 2^n(1+T)}\left|\Delta_{m,n}w\right| + 2\sup_{\tau\in[0,T]}\left|w(\tau)-w_n(\tau)\right|\right].\end{aligned}$$

Now assume that $\alpha \in \left(0,\frac{1}{2}\right)$, and let $p \in [1,\infty)$ be given. Then,

$$\begin{aligned}&\mathbb{E}^{\mathbb{P}}\left[\left(\sup_{\substack{0\le s<t\le T\\ t-s\le 1}} \frac{\left|w(t)-w(s)\right|}{(t-s)^\alpha}\right)^p\right]^{\frac{1}{p}} \\ &\qquad \le 4^{\alpha-\frac{1}{2}} \sum_{n=0}^\infty 2^{n\alpha}\mathbb{E}^{\mathbb{P}}\left[\max_{m\le 2^nT}\left|\Delta_{m,n}w\right|^p\right]^{\frac{1}{p}} \\ &\qquad\quad + 4^{\alpha+\frac{1}{2}} \sum_{n=0}^\infty 2^{n\alpha}\mathbb{E}^{\mathbb{P}}\left[\sup_{\tau\in[0,T]}\left|w(\tau)-w_n(\tau)\right|^p\right]^{\frac{1}{p}}.\end{aligned}$$

Finally, choose $q \in (p,\infty)$ so that $\frac{1}{q'} > 2\alpha$, and observe that

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\left[\max_{m\le 1+2^nT}\left|\Delta_{m,n}w\right|^p\right]^{\frac{1}{p}} &\le \mathbb{E}^{\mathbb{P}}\left[\left(\sum_{m=0}^{2^nT}\left|\Delta_{m,n}w\right|^{2q}\right)^{\frac{p}{2q}}\right]^{\frac{1}{p}} \\ &\le C_q 2^{-\frac{n}{2}}\left(1+2^nT\right)^{\frac{1}{2q}} \le C_q(1+T)^{\frac{1}{2q}}2^{-\frac{n}{2q'}}.\end{aligned}$$

Hence, in conjunction with the estimate obtained in (*), we now see that, for each $\alpha \in \left(0,\frac{1}{2}\right)$ and $p \in [1,\infty)$, there is a $C_p(\alpha) < \infty$ such that

$$\mathbb{E}^{\mu_{\mathbb{R}}}\left[\left(\sup_{0\le s<t\le T} \frac{\left|w(t)-w(s)\right|}{(t-s)^\alpha}\right)^p\right]^{\frac{1}{p}} \le C_p(\alpha)(1+T). \tag{1.6}$$

In § 1.2, we will see that almost no Brownian path is Hölder continuous of any order $\alpha > \frac{1}{2}$. Finding the exact modulus of continuity of a Brownian path requires more effort. This effort was made by P. Lévy, who showed that, $\mu_{\mathbb{R}}$-almost surely,

$$\varlimsup_{\delta\searrow 0} \sup_{0<t-s<\delta} \frac{|w(t)-w(s)|}{\omega(\delta)} = 1, \quad \text{where } \omega(\delta) = \sqrt{2\delta\log\delta^{-1}}.$$

1.1.4. Multi-dimensional Brownian Motion. Thus far we have been looking only at $\mathbb{R}$-valued Brownian paths. However, at least on a cursory level, there is little distinction between the $\mathbb{R}$-valued and $\mathbb{R}^d$-valued settings. That is, let $\mathfrak{W}(\mathbb{R}^d)$ denote the Frechét space of continuous $\mathbf{w} : [0,\infty) \longrightarrow \mathbb{R}^d$ with $\mathbf{w}(0) = \mathbf{0}$. The *standard Wiener measure* $\mu_{\mathbb{R}^d}$ on $\mathfrak{W}(\mathbb{R}^d)$ is the Borel measure with the property that, for each $t \in [0,\infty)$ and $\tau \in (0,\infty)$, $\mathbf{w}(t+\tau) - \mathbf{w}(t)$ is a centered, $\mathbb{R}^d$-valued Gaussian random variable which is independent of

$$\mathcal{B}_t = \sigma\big(\{\mathbf{w}(s) : s \in [0,t]\}\big) \tag{1.7}$$

and has covariance $\tau\mathbf{I}$. In other words, (1.1) gets replaced by

$$\begin{aligned} &\mu_{\mathbb{R}^d}\big(\{\mathbf{w} : \mathbf{w}(t_1) - \mathbf{w}(t_0) \in \Gamma_1, \dots, \mathbf{w}(t_n) - \mathbf{w}(t_{n-1}) \in \Gamma_n\}\big) \\ &\qquad\qquad = \prod_{m=1}^{n} \int_{\Gamma_m} g^{(d)}_{\tau_m}(\eta_m)\, d\eta_m, \end{aligned} \tag{1.8}$$

where (cf. (1.2))

$$g^{(d)}_{\tau}(\boldsymbol{\eta}) = \prod_{k=1}^{d} g_\tau(\eta^k) \quad \text{for } \boldsymbol{\eta} = (\eta^1, \dots, \eta^d) \in \mathbb{R}^d. \tag{1.9}$$

Clearly, for any fixed $\boldsymbol{\xi} \in \mathbf{S}^{d-1}$ (the unit sphere in $\mathbb{R}^d$), the distribution of $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto (\boldsymbol{\xi}, \mathbf{w})_{\mathbb{R}^d} \in \mathfrak{W}(\mathbb{R})$ under $\mu_{\mathbb{R}^d}$ is $\mu_{\mathbb{R}}$. Moreover, by Lemma 1.3, the σ-algebra

$$\sigma\Big(\big\{(\boldsymbol{\xi}, \mathbf{w}(t))_{\mathbb{R}^d} : t \in [0,\infty)\big\}\Big)$$

is $\mu_{\mathbb{R}^d}$-independent of

$$\sigma\Big(\big\{(\boldsymbol{\eta}, \mathbf{w}(t))_{\mathbb{R}^d} : \boldsymbol{\eta} \perp \boldsymbol{\xi} \;\&\; t \in [0,\infty)\big\}\Big).$$

Thus, if $(\mathbf{e}_1, \dots, \mathbf{e}_d)$ is any orthonormal basis in $\mathbb{R}^d$, then $\mu_{\mathbb{R}^d}$ gets mapped (i.e., pushed forward) into $(\mu_{\mathbb{R}})^d$ under the map

$$\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto \big((\mathbf{e}_1, \mathbf{w})_{\mathbb{R}^d}, \dots, (\mathbf{e}_d, \mathbf{w})_{\mathbb{R}^d}\big) \in \mathfrak{W}(\mathbb{R})^d.$$

Conversely, we can realize $\mu_{\mathbb{R}^d}$ by pushing $\mu^d_{\mathbb{R}}$ forward under the map which takes $(w^1, \dots, w^d) \in \mathfrak{W}(\mathbb{R})^d$ into the path $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$ given by

$$\mathbf{w}(t) = \big(w^1(t), \dots, w^d(t)\big), \quad t \in [0,\infty).$$

In other words, given that $\mu_{\mathbb{R}}$ exists, the existence of $\mu_{\mathbb{R}^d}$ presents no difficulty. In addition, as our discussion should have made clear, $\mu_{\mathbb{R}^d}$ is *rotation invariant* in the sense that if O is any orthogonal transformation on $\mathbb{R}^d$ and $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto O\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$ is defined by $(O\mathbf{w})(t) = O\mathbf{w}(t)$, $t \in [0,\infty)$, then, under $\mu_{\mathbb{R}^d}$, $\mathbf{w} \rightsquigarrow O\mathbf{w}$ has the same distribution under $\mu_{\mathbb{R}^d}$ as $\mathbf{w}$ has itself.

1.2 The Infinite Dimensional Sphere and Related Matters

As we have already seen, there are good reasons to consider increments of Brownian paths; and if increments are good, then derivatives ought to be better. However, as Wiener himself was the first to show, *almost no Brownian path is anywhere differentiable*. There are various ways to see this fact, and one of the simplest was provided by A. Dvoretsky (cf. Theorem 4.1.10 in [35]).

1.2.1. Square Variation of Brownian Paths. From our point of view, the nowhere differentiability of Brownian paths is less important than the closely related fact that, for each $T \in (0,\infty)$, $w \upharpoonright [0,T]$ *lives on the infinite dimensional sphere* $\mathbf{S}^\infty(T)$ *of radius* T. More precisely[10]

$$(1.10)\quad \lim_{n\to\infty} \sum_{m=0}^{2^n-1} \Big(w\big((m+1)2^{-n}T\big) - w\big(m2^{-n}T\big)\Big)^2 = T \quad \mu_{\mathbb{R}}\text{-almost surely.}$$

The proof of (1.10) is straight-forward. Namely, set

$$X_{m,n} = \Big(w\big((m+1)2^{-n}T\big) - w\big(m2^{-n}T\big)\Big)^2 - 2^{-n}T,$$

and observe that, for each $n \in \mathbb{N}$, $\{X_{m,n} : 0 \le m < n\}$ is a family of mutually independent, centered random variables with variance $2^{-2n+1}T^2$. Hence, since

$$\sum_{m=0}^{n-1} \Big(w\big((m+1)2^{-n}T\big) - w\big(m2^{-n}T\big)\Big)^2 - T = \sum_{m=0}^{2^n-1} X_{m,n},$$

orthogonality gives

$$\mathbb{E}^{\mu_{\mathbb{R}}}\left[\left(\sum_{m=0}^{2^n-1}\Big(w\big((m+1)2^{-n}T\big) - w\big(m2^{-n}T\big)\Big)^2 - T\right)^2\right] = 2^{-n+1}T^2,$$

and so

$$\sum_{n=0}^{\infty} \mathbb{E}^{\mu_{\mathbb{R}}}\left[\left(\sum_{m=0}^{2^n-1}\Big(w\big((m+1)2^{-n}T\big) - w\big(m2^{-n}T\big)\Big)^2 - T\right)^2\right]^{\frac{1}{2}} < \infty.$$

---

[10] Actually, as explained in Exercise 4.1.11 of [35], even more is true. For related results, see Exercise 2.2.25 in *op. cit.*

Thus (1.10) follows from standard measure theoretic considerations.

Although (1.10) is insufficient to prove that almost no Brownian path is anywhere differentiable, it does show that almost no Brownian path is either Hölder continuous of order strictly greater than $\frac{1}{2}$ or of bounded variation on any interval of positive length. Indeed, if, for some $T \in (0,\infty)$, either one of these properties held for $w \restriction [0,T]$ with positive $\mu_{\mathbb{R}}$-probability, then (because the paths are continuous), with at least the same probability, one would have that

$$\lim_{n\to\infty} \sum_{m=0}^{2^n-1} \Big(w\big((m+1)2^{-n}T\big) - w\big(m2^{-n}T\big)\Big)^2 = 0,$$

which would contradict (1.10).

1.2.2. Paley–Wiener Integrals. Because of (1.10), one cannot afford to be too careless when dealing with integrals of the form

$$\big[\mathcal{I}(f)\big](w) \equiv \int_0^\infty f(t)\,dw(t). \tag{1.11}$$

Indeed, even if $f : [0,\infty) \longrightarrow \mathbb{R}$ is continuous and has compact support, it is not clear what meaning to assign $\big[\mathcal{I}(f)\big](w)$ for $\mu_{\mathbb{R}}$-almost any $w \in \mathfrak{W}(\mathbb{R})$. On the other hand, as long as f has finite variation and compact support, then (cf. Theorem 1.2.7 in [36]) $\big[\mathcal{I}(f)\big](w)$ is well-defined as a Riemann-Stieltjes integral for *every* $w \in \mathfrak{W}(\mathbb{R})$ and, in fact, is equal to the Riemann-Stieltjes integral

$$-\int_{[0,T]} w(t)\,df(t) \quad \text{if } f \restriction [T,\infty) = 0.$$

Hence, under these conditions,

$$\big[\mathcal{I}(f)\big](w) = \lim_{n\to\infty} \sum_{m=0}^{\infty} f\big(mn^{-1}\big)\Big(w\big((m+1)n^{-1}\big) - w\big(mn^{-1}\big)\Big).$$

In particular, by Lemma 1.3, $\mathcal{I}(f)$ under $\mu_{\mathbb{R}}$ is a centered, Gaussian random variable with variance

$$\int_{[0,\infty)} f(t)^2\,dt = \lim_{n\to\infty} n^{-1} \sum_{m=0}^{\infty} f\big(mn^{-1}\big)^2.$$

As noted by Paley and Wiener, this simple observation has an interesting corollary. Namely, it says that

$$\big\|\mathcal{I}(f)\big\|_{L^2(\mu_{\mathbb{R}};\mathbb{R})} = \|f\|_{L^2([0,\infty);\mathbb{R})}$$

for compactly supported f of bounded variation. Thus, there is a unique extension of $f \rightsquigarrow \mathcal{I}(f)$ as an isometry from $L^2([0,\infty);\mathbb{R})$ to $L^2(\mu_{\mathbb{R}};\mathbb{R})$. Note that, in general, $\mathcal{I}(f)$ is defined only up to a $\mu_{\mathbb{R}}$-null set,[11] and so, for example, it is not possible to talk about the whole family

$$\left\{[\mathcal{I}(f)](w) : f \in L^2([0,\infty);\mathbb{R})\right\} \subseteq \mathbb{R}$$

for $\mu_{\mathbb{R}}$-typical $w \in \mathfrak{W}(\mathbb{R})$. Nonetheless, $\{\mathcal{I}(f) : f \in L^2([0,\infty);\mathbb{R})\}$ makes perfectly good sense as a subset of $L^2(\mu_{\mathbb{R}};\mathbb{R})$. In fact, it should be clear that

$$\text{(1.12)} \quad \left\{\mathcal{I}(f) : f \in L^2([0,\infty);\mathbb{R})\right\} = \overline{\operatorname{span}\{w(t) - w(s) : 0 \le s < t\}}^{L^2(\mu_{\mathbb{R}};\mathbb{R})}.$$

1.2.3. Fourier Characterization. In view of the preceding discussion, we know that $\mu_{\mathbb{R}}$ is uniquely characterized (cf. Theorem 4.2.4 in [35]) by the fact that

$$\text{(1.13)} \qquad \mathbb{E}^{\mu_{\mathbb{R}}}\left[e^{\sqrt{-1}\,\mathcal{I}(f)}\right] = \exp\left(-\frac{\|f\|^2_{L^2([0,\infty);\mathbb{R})}}{2}\right), \quad f \in C^1_{\mathrm{c}}((0,\infty);\mathbb{R}).$$

This characterization has several advantages. For example, suppose that $A : \mathfrak{W}(\mathbb{R}) \longrightarrow \mathfrak{W}(\mathbb{R})$ is a measurable linear transformation to which there corresponds a linear isometry $\tilde{A} : L^2([0,\infty);\mathbb{R}) \longrightarrow L^2([0,\infty);\mathbb{R})$ with the property that[12]

$$\tilde{A}f = f \circ A \quad \text{for each } f \in C^1_{\mathrm{c}}((0,\infty);\mathbb{R}).$$

Then $\mu_{\mathbb{R}}$ is invariant under A: $A_*\mu_{\mathbb{R}^d} = \mu_{\mathbb{R}} \circ A^{-1} = \mu_{\mathbb{R}}$. To see this, set $\nu = \mu_{\mathbb{R}} \circ A^{-1}$ and compute:

$$\begin{aligned}\mathbb{E}^{\nu}\left[e^{\sqrt{-1}\,\mathcal{I}(f)}\right] &= \mathbb{E}^{\mu_{\mathbb{R}}}\left[e^{\sqrt{-1}\,\mathcal{I}(\tilde{A}f)}\right] = \exp\left(-\frac{\|\tilde{A}f\|^2_{L^2([0,\infty);\mathbb{R})}}{2}\right) \\ &= \exp\left(-\frac{\|f\|^2_{L^2([0,\infty);\mathbb{R})}}{2}\right).\end{aligned}$$

A rather trivial source of such maps[13] is provided by the group of *Wiener scaling* transformations $\{S_\alpha : \alpha > 0\}$ given by $[S_\alpha w](t) = \alpha^{-\frac{1}{2}} w(\alpha t)$, $t \in [0,\infty)$. Clearly, the associated $\tilde{S}_\alpha$ is the $L^2([0,\infty);\mathbb{R})$-isometry given by $[\tilde{S}_\alpha f](t) = \alpha^{-\frac{1}{2}} f(\alpha^{-1}t)$.

1.2.4. Extension to Higher Dimensions. Of course, all the preceding considerations extend, without difficulty, to Brownian paths in $\mathbb{R}^d$. The only

[11] If ν is a measure, then we say the Γ is a ν-null set if $\nu(\Gamma) = 0$.

[12] We use the subscript "c" to mean *compactly supported*. Thus, here the space is continuously differentiable $\mathbb{R}$-valued functions with compact support on $[0,\infty)$.

[13] There are many other examples. One of the more intriguing is the Wiener time-inversion transformation discussed in part (iii) of Exercise 4.2.39 in [35].

change is that one has to replace $f \in L^2([0,\infty);\mathbb{R})$ by $\mathbf{f} \in L^2([0,\infty);\mathbb{R}^d)$ and take

$$(1.14) \qquad [\mathcal{I}(\mathbf{f})](\mathbf{w}) = \sum_{k=1}^{d} [\mathcal{I}(f^k)](w^k),$$

where $\{f_k\}_1^d$ and $\{w_k\}_1^d$ are the coordinates of $\mathbf{f}$ and $\mathbf{w}$ with respect to some orthonormal basis in $\mathbb{R}^d$. Thus, for example, $\mu_{\mathbb{R}^d}$ is characterized on $\mathfrak{W}(\mathbb{R}^d)$ by

$$(1.15) \qquad \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[e^{\sqrt{-1}\,\mathcal{I}(\mathbf{f})}\right] = \exp\left(-\frac{\|\mathbf{f}\|^2_{L^2([0,\infty);\mathbb{R}^d)}}{2}\right), \quad \mathbf{f} \in C^1_c\big((0,\infty);\mathbb{R}^d\big);$$

and the obvious statement of Wiener scaling invariance holds for $\mu_{\mathbb{R}^d}$ just as did for $\mu_{\mathbb{R}}$. In fact, given an orthonormal basis in $\mathbb{R}^d$, $\mu_{\mathbb{R}^d}$ is invariant even when one applies different scale changes in different coordinates.

1.2.5. The Cameron–Martin Formula. One of the complicating features of measures on infinite dimensional spaces is that they tend to be inimical to one another. Thus, for example, even apparently benign transformations like translation usually take a given measure into one which is singular. To wit, given $\mathbf{h} \in \mathfrak{W}(\mathbb{R}^d)$, consider the translation $T_{\mathbf{h}} : \mathfrak{W}(\mathbb{R}^d) \longrightarrow \mathfrak{W}(\mathbb{R}^d)$ given by $T_{\mathbf{h}}\mathbf{w} = \mathbf{h} + \mathbf{w}$. Is $\mu_{\mathbb{R}^d} \circ T_{\mathbf{h}}^{-1} \ll \mu_{\mathbb{R}^d}$? The answer is that "it depends," but certainly "not in general." Indeed, take $d = 1$ and $h \in \mathfrak{W}$. Then (cf. (1.10)), for $\mu_{\mathbb{R}}$-almost every $w \in \mathfrak{W}(\mathbb{R})$:

$$\begin{aligned}
&\varlimsup_{n\to\infty} \sum_{m=1}^{2^n-1} \Big(T_h w\big((m+1)2^{-n}\big) - T_h w\big(m2^{-n}\big)\Big)^2 \\
&\geq 1 + \varlimsup_{n\to\infty} \sum_{m=1}^{2^n-1} \Big(h\big((m+1)2^{-n}\big) - h\big(m2^{-n}\big)\Big)^2 \\
&\quad + 2 \varliminf_{n\to\infty} \sum_{m=1}^{2^n-1} \Big(w\big((m+1)2^{-n}\big) - w\big(m2^{-n}\big)\Big)\Big(h\big((m+1)2^{-n}\big) - h\big(m2^{-n}\big)\Big) \\
&\geq 1 + \underline{\mathrm{var}}_2(h)
\end{aligned}$$

if

$$\varliminf_{n\to\infty} \sum_{m=1}^{2^n-1} \Big(w\big((m+1)2^{-n}\big) - w\big(m2^{-n}\big)\Big)\Big(h\big((m+1)2^{-n}\big) - h\big(m2^{-n}\big)\Big) \geq 0$$

and

$$\underline{\mathrm{var}}_2(h) \equiv \varliminf_{n\to\infty} \sum_{m=1}^{2^n-1} \Big(h\big((m+1)2^{-n}\big) - h\big(m2^{-n}\big)\Big)^2.$$

But clearly, $w^{\leadsto} - w$ has the same distribution under $\mu_{\mathbb{R}}$ as w, and therefore

$$\underline{\mathrm{var}}_2(h) > 0 \implies \mu_{\mathbb{R}}\Big(\big\{w : \underline{\mathrm{var}}_2(T_h w) > 1\big\}\Big) \geq \frac{1}{2}.$$

In particular, after comparing this to (1.10), we conclude that

$$\underline{\mathrm{var}}_2(h) > 0 \implies \mu_{\mathbb{R}} \circ T_h^{-1} \not\ll \mu_{\mathbb{R}}.$$

As the preceding example makes explicit, $\mu_{\mathbb{R}^d}$ is far from being even quasi-invariant under translation. In fact, together with (1.10), it shows that $\mu_{\mathbb{R}^d} \circ T_{\mathbf{h}}^{-1} \not\ll \mu_{\mathbb{R}^d}$ for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{h} \in \mathfrak{W}(\mathbb{R}^d)$. However, the reasoning just given is far too crude to reveal the whole story. Namely, as I. Segal showed (cf. Exercise 5.2.38 in [35]), $\mu_{\mathbb{R}^d} \circ T_{\mathbf{h}}^{-1} \ll \mu_{\mathbb{R}^d}$ only if $\mathbf{h} \in H^1(\mathbb{R}^d)$, where $H^1(\mathbb{R}^d)$ denotes the space of absolutely continuous $\mathbf{h} \in \mathfrak{W}(\mathbb{R}^d)$ with[14]

$$\|\mathbf{h}\|_{H^1(\mathbb{R}^d)} \equiv \|\dot{\mathbf{h}}\|_{L^2([0,\infty);\mathbb{R}^d)} < \infty.$$

For comparison purposes it is helpful to know that $\mathbf{h} \in \mathfrak{W}(\mathbb{R}^d)$ is an element of $H^{(1)}(\mathbb{R}^d)$ if and only if[15]

$$\sup_{n\in\mathbb{N}} 2^n \sum_{m=0}^{\infty} \big|\mathbf{h}\big((m+1)2^{-n}\big) - \mathbf{h}\big(m2^{-n}\big)\big|^2 < \infty.$$

In particular, $h \in H^{(1)}(\mathbb{R})$ is a much stronger statement than $\underline{\mathrm{var}_2}(h) = 0$.

At the same time, Segal, extending an earlier result due R. Cameron and T. Martin, showed that when $\mathbf{h} \in H^1(\mathbb{R}^d)$:

$$\begin{aligned}(1.16)\qquad &\big(\mu_{\mathbb{R}^d} \circ T_{\mathbf{h}}^{-1}\big)(d\mathbf{w}) = R_{\mathbf{h}}(\mathbf{w})\,\mu_{\mathbb{R}^d}(d\mathbf{w}) \\ &\text{where } R_{\mathbf{h}} = \exp\Big(\mathcal{I}(\dot{\mathbf{h}}) - \tfrac{1}{2}\|\mathbf{h}\|^2_{H^1(\mathbb{R}^d)}\Big),\end{aligned}$$

a fact which is called the *Cameron–Martin formula.* Segal's conclusions are dramatic: they say that, for any $\mathbf{h} \in \mathfrak{W}(\mathbb{R}^d)$, $\mu_{\mathbb{R}^d} \circ T_{\mathbf{h}}^{-1}$ is either singular to $\mu_{\mathbb{R}^d}$ or equivalent (i.e., mutually absolutely continuous) with $\mu_{\mathbb{R}^d}$, according to whether $\mathbf{h}$ is not or is an element of $H^1(\mathbb{R}^d)$. Further, when $\mathbf{h} \in H^1(\mathbb{R}^d)$. the Radon–Nikodym derivative is remarkably well-behaved. For instance, because $\mathcal{I}(\dot{\mathbf{h}})$ under $\mu_{\mathbb{R}^d}$ is a centered, Gaussian random variable with variance $\|\mathbf{h}\|^2_{H^1(\mathbb{R}^d)}$,

$$(1.17)\qquad \big\|R_{\mathbf{h}}\big\|_{L^p(\mu_{\mathbb{R}^d};\mathbb{R})} = \exp\left(\frac{p-1}{2}\|\mathbf{h}\|^2_{H^1(\mathbb{R}^d)}\right), \quad p \in (0,\infty).$$

[14] Below, and elsewhere, we use a "dot" to denote derivatives with respect to time.

[15] The verification is quite easy. Namely, for each $n \in \mathbb{N}$, let Π_n denote orthogonal projection onto the subspace of $H^{(1)}(\mathbb{R})$ consisting of elements which are linear on each dyadic interval $[m2^{-n},(m+1)2^{-n}]$, and note that $\|\Pi_n h\|^2_{H^{(1)}(\mathbb{R})}$ is equal $2^n \sum_m |h((m+1)2^{-n}) - h(m2^{-n})|^2$.

Given (1.15), the proof of (1.16) is surprisingly easy. To see this, first observe that, by an essentially trivial analytic continuation from the case when ξ and η are real,

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\exp\big(\xi\mathcal{I}(\mathbf{f})+\eta\mathcal{I}(\mathbf{g})\big)\Big] = \exp\left(\frac{\xi^2\|\mathbf{f}\|^2_{L^2([0,\infty);\mathbb{R}^d)}}{2}+\xi\eta\big(\mathbf{f},\mathbf{g}\big)_{L^2([0,\infty);\mathbb{R}^d)}+\frac{\eta^2\|\mathbf{g}\|^2_{L^2([0,\infty);\mathbb{R}^d)}}{2}\right)$$

for all $(\xi,\eta)\in\mathbb{C}^2$ and $(\mathbf{f},\mathbf{g})\in L^2([0,\infty);\mathbb{R}^d)^2$. Now let $\mathbf{h}$ be a smooth element of $H^1(\mathbb{R}^d)$ with compact support. Then, for any $\mathbf{f}\in C^1_{\mathrm{c}}((0,\infty);\mathbb{R}^d)$,

$$\big[\mathcal{I}(\mathbf{f})\big](\mathbf{w}+\mathbf{h}) = \big[\mathcal{I}(\mathbf{f})\big](\mathbf{w}) + \big(\mathbf{f},\dot{\mathbf{h}}\big)_{L^2([0,\infty);\mathbb{R}^d)}$$

while

$$\big[\mathcal{I}(\dot{\mathbf{h}})\big](\mathbf{w}+\mathbf{h}) = \big[\mathcal{I}(\dot{\mathbf{h}})\big](\mathbf{w}) + \|\mathbf{h}\|^2_{H^1(\mathbb{R}^d)}.$$

Thus, if

$$\nu(d\mathbf{w}) = \frac{1}{R_{\mathbf{h}}(\mathbf{w})}\,\big(\mu_{\mathbb{R}^d}\circ T_{\mathbf{h}}^{-1}\big)(d\mathbf{w}),$$

then

$$\begin{aligned}
&\mathbb{E}^{\nu}\Big[e^{\sqrt{-1}\,\mathcal{I}(\mathbf{f})}\Big]\\
&= \exp\left(-\frac{\|\mathbf{h}\|^2_{H^1(\mathbb{R}^d)}}{2}+\sqrt{-1}\,\big(\mathbf{f},\dot{\mathbf{h}}\big)_{L^2([0,\infty);\mathbb{R}^d)}\right)\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\exp\Big(\sqrt{-1}\,\mathcal{I}(\mathbf{f})-\mathcal{I}(\dot{\mathbf{h}})\Big)\Big]\\
&= \exp\left(-\frac{\|\mathbf{f}\|^2_{L^2([0,\infty);\mathbb{R}^d)}}{2}\right).
\end{aligned}$$

Hence, by the characterization given in (1.15), $\nu=\mu_{\mathbb{R}^d}$. Equivalently, (1.16) holds when $\mathbf{h}\in H^1(\mathbb{R}^d)$ is smooth and compactly supported. Finally, the extension to general $\mathbf{h}\in H^1(\mathbb{R}^d)$ now follows by applying an easy approximation procedure and using (1.17) to take care of uniform integrability questions.

1.2.6. Integration by Parts. Just as the central role played by Lebesgue measure in finite dimensional classical analysis derives from its translation invariance, the importance of Wiener measure in infinite dimensional analysis rests on the fact that $\mu_{\mathbb{R}^d}$ is as translation invariant as it reasonable to hope any measure on an infinite dimensional space can be. In particular, one can start doing differential calculus on the basis of (1.16). Namely, given any $F\in\bigcup_{p\in(1,\infty)}L^p(\mu_{\mathbb{R}^d};\mathbb{R})$ and $\mathbf{h}\in H^1(\mathbb{R}^d)$, (1.16) and (1.17) combine to guarantee that $F\circ T_{\mathbf{h}}$ is a well-defined element $L^1(\mu_{\mathbb{R}^d};\mathbb{R})$ and

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\big[F\circ T_{\mathbf{h}}\big] = \mathbb{E}^{\mu_{\mathbb{R}^d}}\big[F\,R_{\mathbf{h}}\big].$$

Hence, after replacing $\mathbf{h}$ by $s\mathbf{h}$ and considering difference quotients, one can easily use uniform integrability considerations to justify

$$\begin{aligned}\frac{d}{ds}\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[F\circ T_{s\mathbf{h}}\right]\Big|_{s=0} &= \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[F\mathcal{I}(\dot{\mathbf{h}})\right]\\ \text{for all } F\in \bigcup_{p\in(1,\infty)} L^p(\mu_{\mathbb{R}^d};\mathbb{R}) &\text{ and } \mathbf{h}\in H^1(\mathbb{R}^d).\end{aligned} \tag{1.18}$$

In that it leads immediately to an integration by parts formula, it is clear that (1.18) provides the foundation on which one can start doing analysis. That is, under appropriate conditions on F, (1.18) yields formulae like

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[D_{\mathbf{h}}F\right] = \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[F\mathcal{I}(\dot{\mathbf{h}})\right], \tag{1.19}$$

where $D_{\mathbf{h}}F$ denotes the derivative of F in the direction $\mathbf{h}$.

1.3 Feynman's Picture of Wiener Measure

The facts just described encourage the completely false but intuitively appealing "description" of Wiener measure $\mu_{\mathbb{R}^d}$ as *the standard centered Gaussian measure on* $H^1(\mathbb{R}^d)$.

To be more precise about this inherently flawed idea, notice that $H^1(\mathbb{R}^d)$ is a separable Hilbert space which sits inside $\mathfrak{W}(\mathbb{R}^d)$ as a Borel subset. Moreover, as we already observed in § 1.2.5, (1.10) tells us that $H^1(\mathbb{R}^d)$ is a $\mu_{\mathbb{R}^d}$-null set. Thus, from the outset, there is something very wrong with any statement which seems to imply that $\mu_{\mathbb{R}^d}\left(H^1(\mathbb{R}^d)\right) = 1$. Nonetheless, suppose that we agree to suspend our mathematically mandated disbelief and ignore this "technical quibble." That is, we will proceed as if $\mu_{\mathbb{R}^d}\left(H^1(\mathbb{R}^d)\right) = 1$. In fact, we are about to commit much more egregious sins. Namely, we will ignore the fact that $H^1(\mathbb{R}^d)$ is infinite dimensional and will pretend that it admits a translation invariant Borel measure $\lambda_{H^1(\mathbb{R}^d)}$ which assigns mass one to any unit cube in $H^1(\mathbb{R}^d)$. Finally, having already thoroughly committed ourselves to this line of iniquity, we will go one step further and pretend that $\mu_{\mathbb{R}^d} \ll \lambda_{H^1(\mathbb{R}^d)}$.

It is hard to imagine that anything interesting or worthwhile could come out of all these ridiculous concessions. Nonetheless, something does. In the first place, because we are now pretending that $\mu_{\mathbb{R}^d}$ is supported by $H^1(\mathbb{R}^d)$, we must also pretend that

$$\left[\mathcal{I}(\dot{\mathbf{h}})\right](\mathbf{w}) = \left(\mathbf{w},\mathbf{h}\right)_{H^1([0,\infty);\mathbb{R}^d)}.$$

At the same time, we can re-interpret (1.19) as saying that (in the sense of a L. Schwartz-type distribution theory for $H^1(\mathbb{R}^d)$)

$$D_{\mathbf{h}}g = -\mathcal{I}(\dot{\mathbf{h}})g, \quad \mathbf{h}\in H^1(\mathbb{R}^d),$$

where g is the supposed "Radon–Nikodym derivative" of $\mu_{\mathbb{R}^d}$ with respect to $\lambda_{H^1(\mathbb{R}^d)}$. Hence, after combining these two, we arrive at

$$\left[D_{\mathbf{h}}\log g\right](\mathbf{w}) = -\left(\mathbf{w},\mathbf{h}\right)_{H^1(\mathbb{R}^d)}, \quad (\mathbf{w},\mathbf{h})\in H^1(\mathbb{R}^d)^2,$$

which leads immediately to the conclusion

$$\frac{d\mu_{\mathbb{R}^d}}{d\lambda_{H^1(\mathbb{R}^d)}}(\mathbf{w}) = \frac{1}{Z_{H^1(\mathbb{R}^d)}} \exp\left(-\frac{\|\mathbf{w}\|^2_{H^1(\mathbb{R}^d)}}{2}\right), \tag{1.20}$$

with the "constant" $Z_{H^1(\mathbb{R}^d)} = \left(\sqrt{2\pi}\right)^{\dim(H^1(\mathbb{R}^d))}$ is determined by the fact that $\mu_{\mathbb{R}^d}$ has total mass 1.

For whatever it is worth (which turns out to be quite a lot), the "equation" in (1.20) is more or less the one which R. Feynman would have written down had he bothered with pathspace integrals as *trivial* as those which can be rationalized in terms of Wiener measure. The remarkable fact about (1.20) is that, in spite of the absence on the right hand side of even one component which makes rigorous mathematical sense, essentially everything which one can predict on the basis (1.20) turns out to be true, at least after appropriate interpretation. For example, if it makes any sense, (1.20) would seem to be saying what we said at the beginning of this discussion: *$\mu_{\mathbb{R}^d}$ is the standard, centered Gaussian measure on $H^1(\mathbb{R}^d)$.* If such a quantity actually existed, it would certainly have the property that

$$\int_{H^1(\mathbb{R}^d)} \exp\left(\sqrt{-1}\,(\mathbf{w},\mathbf{h})_{H^1(\mathbb{R}^d)}\right)\mu(d\mathbf{w}) = \exp\left(-\frac{\|\mathbf{h}\|^2_{H^1(\mathbb{R}^d)}}{2}\right), \quad \mathbf{h} \in H^1(\mathbb{R}^d).$$

The secret for passing from non-sense like the preceding to rigorous mathematical formulae lies in the systematic replacement of all the expressions which do not make sense by ones which do. In the present example, there is no problem with the right hand side, but the left hand side seems to require that $\mu_{\mathbb{R}^d}$ live on $H^1(\mathbb{R}^d)$. On the other hand, we have already discovered that the "correct" interpretation of $\mathbf{w}\rightsquigarrow(\mathbf{w},\mathbf{h})_{H^1(\mathbb{R}^d)}$ is $\mathcal{I}(\dot{\mathbf{h}})$; and, as soon as we adopt this interpretation, the preceding is seen to become (1.15) with $\mathbf{f} = \dot{\mathbf{h}}$. Similarly, starting from (1.20), one would predict that

$$\frac{d\mu_{\mathbb{R}^d}\circ T_{\mathbf{h}}^{-1}}{d\lambda_{H^1(\mathbb{R}^d)}}(\mathbf{w}) = \frac{1}{Z_{H^1(\mathbb{R}^d)}} \exp\left(-\frac{\|\mathbf{w}-\mathbf{h}\|^2_{H^1(\mathbb{R}^d)}}{2}\right),$$

which would say that

$$\frac{d\mu_{\mathbb{R}^d}\circ T_{\mathbf{h}}^{-1}}{d\mu_{\mathbb{R}^d}}(\mathbf{w}) = \exp\left((\mathbf{w},\mathbf{h})_{H^1(\mathbb{R}^d)} - \frac{\|\mathbf{h}\|^2_{H^1(\mathbb{R}^d)}}{2}\right).$$

Hence, after again replacing $\mathbf{w}\rightsquigarrow(\mathbf{w},\mathbf{h})_{H^1(\mathbb{R}^d)}$ by $\mathcal{I}(\dot{\mathbf{h}})$, we see that this prediction is precisely (1.16).

1.3.1. Rescaling Feynman's Picture. Given $T \in (0,\infty)$ and $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$, define $\mathbf{w}^{(T)} \in \mathfrak{W}(\mathbb{R}^d)$ so that $\mathbf{w}^{(T)}(t) = \mathbf{w}(Tt)$ for $t \in [0,\infty)$ and let $\mu^{(T)}_{\mathbb{R}^d}$

denote the distribution of $\mathbf{w} \rightsquigarrow \mathbf{w}^{(T)}$ under $\mu_{\mathbb{R}^d}$. By (1.10), we know that $\mu_{\mathbb{R}^d}^{(T)}$ is singular to $\mu_{\mathbb{R}^d}$. On the other hand, by Wiener scaling invariance, we know that $\mu_{\mathbb{R}^d}^{(T)}$ is the distribution of $\mathbf{w} \rightsquigarrow T^{\frac{1}{2}}\mathbf{w}$ under $\mu_{\mathbb{R}^d}$. Thus if F is a bounded, continuous function on $\mathfrak{W}(\mathbb{R}^d)$, then

$$\mathbb{E}^{\mu_{\mathbb{R}^d}^{(T)}}\big[F\big] = \mathbb{E}^{\mu_{\mathbb{R}^d}}\big[F_T\big] \quad \text{where } F_T(\mathbf{w}) \equiv F\big(T^{\frac{1}{2}}\mathbf{w}\big).$$

We now want to use this observation to develop Feynman's picture of $\mu_{\mathbb{R}^d}^{(T)}$. Namely, according to (1.20), the preceding should be written as

$$\mathbb{E}^{\mu_{\mathbb{R}^d}^{(T)}}\big[F\big] = \frac{1}{Z_{H^1(\mathbb{R}^d)}} \int_{H^1(\mathbb{R}^d)} F_T(\mathbf{w}) \exp\left(-\frac{\|\mathbf{w}\|^2_{H^1(\mathbb{R}^d)}}{2}\right) \lambda_{H^1(\mathbb{R}^d)}(d\mathbf{w}),$$

and, after a change of variable, this becomes

$$\mathbb{E}^{\mu_{\mathbb{R}^d}^{(T)}}\big[F\big] = \frac{1}{Z_{H^1(\mathbb{R}^d)}(T)} \int_{H^1(\mathbb{R}^d)} F(\mathbf{w}) \exp\left(-\frac{\|\mathbf{w}\|^2_{H^1(\mathbb{R}^d)}}{2T}\right) \lambda_{H^1(\mathbb{R}^d)}(d\mathbf{w}),$$

where "$Z_{H^1(\mathbb{R}^d)}(T) = \big(\sqrt{2\pi T}\big)^{\dim(H^1(\mathbb{R}^d))}$". In other words, the Feynman picture of $\mu_{\mathbb{R}^d}^{(T)}$ is that

$$\text{(1.21)} \qquad \frac{d\mu_{\mathbb{R}^d}^{(T)}}{d\lambda_{H^1(\mathbb{R}^d)}}(\mathbf{w}) = \frac{1}{Z_{H^1(\mathbb{R}^d)}(T)} \exp\left(-\frac{\|\mathbf{w}\|^2_{H^1(\mathbb{R}^d)}}{2T}\right).$$

The virtue of (1.21) is that it highlights an important property of Brownian paths. Namely, given an event $A \in \mathcal{B}_{\mathfrak{W}(\mathbb{R}^d)}$, set

$$I(A) = \inf\left\{\frac{\|\mathbf{h}\|^2_{H^1(\mathbb{R}^d)}}{2} : \mathbf{h} \in A\right\}.$$

Then a naive look at (1.21) might make one say that, as $T \searrow 0$, the major contribution to $\mu_{\mathbb{R}^d}^{(T)}(A)$ will be made by paths $\mathbf{w} \in A$ for which $\|\mathbf{w}\|_{H^1(\mathbb{R}^d)}$ is nearly $I(A)$. Of course, since $\mu_{\mathbb{R}^d}^{(T)}$-almost no $\mathbf{w} \in H^1(\mathbb{R}^d)$, this naive interpretation needs amendment: one needs to take into account the topology of $\mathfrak{W}(\mathbb{R}^d)$ and consider paths $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$ which lie in a neighborhood of an $\mathbf{h} \in A$ with $\|\mathbf{h}\|_{H^1(\mathbb{R}^d)} \sim I(A)$.[16] Nonetheless, as usual, the idea conveyed by Feynman's picture is both clear and (essentially) correct.

[16] The basic mathematically correct statement of this sort is due to Schilder (cf. Theorem 1.3.27 in [7]).

1.4 Wiener Measure, the Laplacian, and Martingales

In this concluding section, we will describe yet another way in which to look at Wiener measure. This one, as distinguished from the one given by Feynman's picture, is easy to turn into rigorous mathematics. In fact, it will be the description on which we will base our description of more general measures on pathspace.

Up to now, everything which we have said about Wiener measure depends on the fact that Wiener measure is a Gaussian measure. Thus, much of what we have said would transfer, *mutatis mutandis*, to any other Gaussian measure on an infinite dimensional vector space. By contrast, the description which we about to give relies entirely on the fact that the space on which our Wiener measure lives is a pathspace. In particular, it will emphasize the property that $\mathfrak{W}(\mathbb{R}^d)$ has a natural time-ordering and will de-emphasize the property that $\mathfrak{W}(\mathbb{R}^d)$ is linear. As a consequence, this new description will lend itself better to the analysis of measures on pathspaces for which there is no linear structure, especially, spaces of paths with values in a differentiable manifold.

1.4.1. A Preliminary Manipulation. We begin by recalling that $\mu_{\mathbb{R}^d}$ is the unique Borel probability measure on $\mathfrak{W}(\mathbb{R}^d)$ with the property that, for each $t \in [0,\infty)$ and $\tau \in (0,\infty)$, $\mathbf{w} \rightsquigarrow \mathbf{w}(t+\tau) - \mathbf{w}(t)$ under $\mu_{\mathbb{R}^d}$ is a centered, $\mathbb{R}^d$-valued Gaussian random variable which is independent of the σ-algebra $\mathcal{B}_t$ in (1.7) and has covariance $\tau\mathbf{I}$. Equivalently, for any bounded, $\mathcal{B}_{\mathbb{R}^d} \times \mathcal{B}_t$-measurable $F : \mathbb{R}^d \times \mathfrak{W}(\mathbb{R}^d) \longrightarrow \mathbb{R}$ and $A \in \mathcal{B}_t$, (cf. (1.9))

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[F\big(\mathbf{w}(t+\tau) - \mathbf{w}(t), \mathbf{w}\big)\Big] = \mathbb{E}^{\mu_{\mathbb{R}^d}}\big[\tilde{F}(\mathbf{w})\big]$$
$$\text{where } \tilde{F}(\mathbf{w}) \equiv \int_{\mathbb{R}^d} F(\boldsymbol{\eta}, \mathbf{w}) g_\tau^{(d)}(\boldsymbol{\eta})\, d\boldsymbol{\eta}.$$

In particular, if $f : C^2_{\mathrm{b}}(\mathbb{R}^d;\mathbb{R})$, $A \in \mathcal{B}_t$, and we take[17]

$$F(\boldsymbol{\xi}, \mathbf{w}) \equiv f\big(\boldsymbol{\xi} + \mathbf{w}(t)\big)\mathbf{1}_A(\mathbf{w})$$

in the preceding, then we find that

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[f\big(\mathbf{w}(t+\tau)\big),\, A\Big] = \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\int_{\mathbb{R}^d} f\big(\mathbf{w}(t) + \boldsymbol{\eta}\big) g_\tau^{(d)}(\boldsymbol{\eta})\, d\boldsymbol{\eta},\, A\right]$$

To carry out the next step, note that[18]

$$\partial_\tau g_\tau^{(d)}(\boldsymbol{\eta}) \equiv \frac{\partial}{\partial\tau} g_\tau^{(d)}(\boldsymbol{\eta}) = \tfrac{1}{2}\big[\Delta_{\mathbb{R}^d} g_\tau^{(d)}\big](\boldsymbol{\eta}) \quad \text{in } (0,\infty) \times \mathbb{R}^d,$$

[17] Below, and elsewhere, we use $\mathbf{1}_\Gamma$ to denote the indicator function (also known as characteristic function in the analytic literature) of the set Γ.

[18] $\Delta_{\mathbb{R}^d}$ is used to denote the analyst's (i.e., the non-positive one) standard Laplace operator on $C^2(\mathbb{R}^d;\mathbb{R})$.

and therefore (after elementary integration by parts)

$$\frac{d}{d\tau}\int_{\mathbb{R}^d} f\big(\mathbf{w}(t)+\boldsymbol{\eta}\big)g_\tau^{(d)}(\boldsymbol{\eta})\,d\boldsymbol{\eta} = \int_{\mathbb{R}^d} \tfrac{1}{2}\big[\Delta_{\mathbb{R}^d} f\big]\big(\mathbf{w}(t)+\boldsymbol{\eta}\big)g_\tau^{(d)}(\boldsymbol{\eta})\,d\boldsymbol{\eta}.$$

Hence,

$$\begin{aligned}&\frac{d}{d\tau}\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[f\big(\mathbf{w}(t+\tau)\big),\,A\Big]\\&\quad=\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\int_{\mathbb{R}^d}\tfrac{1}{2}\big[\Delta_{\mathbb{R}^d} f\big]\big(\mathbf{w}(t)+\boldsymbol{\eta}\big)g_\tau^{(d)}(\boldsymbol{\eta})\,d\boldsymbol{\eta},\,A\right]\\&\quad=\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\tfrac{1}{2}\big[\Delta_{\mathbb{R}^d} f\big]\big(\mathbf{w}(t+\tau)\big),\,A\Big],\end{aligned}$$

which, by the Fundamental Theorem of Calculus and Fubini's Theorem, means that

$$\begin{aligned}&\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[f\big(\mathbf{w}(t+\tau)\big)-f\big(\mathbf{w}(t)\big),\,A\Big]\\&\qquad=\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\int_t^{t+\tau}\tfrac{1}{2}\big[\Delta_{\mathbb{R}^d} f\big]\big(\mathbf{w}(s)\big)\,ds,\,A\right].\end{aligned}\tag{1.22}$$

Before going further, one should notice that (1.22) for every $\tau>0$, $t\in[0,\infty)$, $A\in\mathcal{B}_t$, and $f\in C_{\mathrm{b}}^\infty(\mathbb{R}^d;\mathbb{R})$ is equivalent to the defining property of $\mu_{\mathbb{R}^d}$ with which we began this discussion. Indeed, suppose that ν is any Borel probability measure on $\mathfrak{W}(\mathbb{R}^d)$ for which (1.22) holds when $\mu_{\mathbb{R}^d}$ is replaced by ν. To see that $\nu=\mu_{\mathbb{R}^d}$, first note that there is no difficulty in checking that (1.22) self-extends to $f\in C_{\mathrm{b}}^\infty(\mathbb{R}^d;\mathbb{C})$. In particular, given any $\boldsymbol{\theta}\in\mathbb{R}^d$, take

$$f(\boldsymbol{\xi})=\exp\Big(\sqrt{-1}\,(\boldsymbol{\theta},\boldsymbol{\xi})_{\mathbb{R}^d}\Big)$$

in (1.22), and conclude first that

$$\frac{du}{d\tau}(\tau)=-\frac{|\boldsymbol{\theta}|^2u(\tau)}{2}\quad\text{where } u(\tau)\equiv\mathbb{E}^\nu\Big[\exp\Big(\sqrt{-1}\,\big(\boldsymbol{\theta},\mathbf{w}(\tau+1)\big)_{\mathbb{R}^d}\Big),\,A\Big]$$

and then that

$$\begin{aligned}&\mathbb{E}^\nu\Big[\exp\Big(\sqrt{-1}\,\big(\boldsymbol{\theta},\mathbf{w}(t+\tau)\big)_{\mathbb{R}^d}\Big),\,A\Big]\\&\qquad=\exp\Big(-\tfrac{\tau|\boldsymbol{\theta}|^2}{2}\Big)\mathbb{E}^\nu\Big[\exp\Big(\sqrt{-1}\,\big(\boldsymbol{\theta},\mathbf{w}(t)\big)_{\mathbb{R}^d}\Big),\,A\Big].\end{aligned}$$

Because the preceding holds for every $A\in\mathcal{B}_t$, elementary measure theory says that

$$\begin{aligned}&\mathbb{E}^\nu\Big[\exp\Big(\sqrt{-1}\,\big(\boldsymbol{\theta},\mathbf{w}(t+\tau)\big)_{\mathbb{R}^d}\Big)F(\mathbf{w})\Big]\\&\qquad=\exp\Big(-\tfrac{\tau|\boldsymbol{\theta}|^2}{2}\Big)\mathbb{E}^\nu\Big[\exp\Big(\sqrt{-1}\,\big(\boldsymbol{\theta},\mathbf{w}(t)\big)_{\mathbb{R}^d}\Big)F(\mathbf{w})\Big]\end{aligned}$$

for every bounded, $\mathcal{B}_t$-measurable $F : \mathfrak{W}(\mathbb{R}^d) \longrightarrow \mathbb{C}$. Thus, by taking

$$F(\mathbf{w}) = \exp\Big(-\sqrt{-1}\,\big(\boldsymbol{\theta}, \mathbf{w}(t)\big)_{\mathbb{R}^d}\Big)\mathbf{1}_A(\mathbf{w}),$$

we conclude that

$$\mathbb{E}^\nu\Big[\exp\Big(\sqrt{-1}\,\big(\boldsymbol{\theta}, \mathbf{w}(t+\tau) - \mathbf{w}(t)\big)_{\mathbb{R}^d}\Big),\, A\Big] = \exp\Big(-\tfrac{\tau|\boldsymbol{\theta}|^2}{2}\Big)\mu_{\mathbb{R}^d}(A).$$

Finally, because it holds for all $\boldsymbol{\theta} \in \mathbb{R}^d$ and $A \in \mathcal{B}_t$, the preceding is equivalent to the statement that $\mathbf{w} \rightsquigarrow \mathbf{w}(t+\tau) - \mathbf{w}(t)$ under ν is a centered, $\mathbb{R}^d$-valued, Gaussian random variable which is independent of $\mathcal{B}_t$ and has covariance $\tau\mathbf{I}$.

Notice that these considerations have bought us potential freedom from our dependence on the linear structure $\mathfrak{W}(\mathbb{R}^d)$. Indeed, we have in (1.22) a characterization of $\mu_{\mathbb{R}^d}$ which does not make explicit reference to its Gaussian character. Of course, the reference to Gaussian measures is there, but it is there in its differential version, which is the role of the operator $\frac{1}{2}\Delta_{\mathbb{R}^d}$ in (1.22).

1.4.2. Reinterpretation. The next step is nothing more than a reinterpretation of (1.22). Namely, when put in the language of conditional expectation values,[19] (1.22) says that, given $\mathcal{B}_t$, the $\mu_{\mathbb{R}^d}$-conditional expectation values of

$$f\big(\mathbf{w}(t+\tau)\big) - f\big(\mathbf{w}(t)\big) \quad \text{and} \quad \int_t^{t+\tau} \frac{1}{2}\big[\Delta_{\mathbb{R}^d} f\big](\mathbf{w}(s))\,ds$$

coincide. Equivalently, if

$$M_f^{\Delta_{\mathbb{R}^d}}(t, \mathbf{w}) \equiv f\big(\mathbf{w}(t)\big) - \int_0^t \frac{1}{2}\big[\Delta_{\mathbb{R}^d} f\big](\mathbf{w}(s))\,ds,$$

then, because $\mathbf{w} \rightsquigarrow M_f(t, \mathbf{w})$ is $\mathcal{B}_t$-measurable, we see that

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\big[M_f^{\Delta_{\mathbb{R}^d}}(t+\tau)\,\big|\,\mathcal{B}_t\big](\mathbf{w}) = M_f^{\Delta_{\mathbb{R}^d}}(t, \mathbf{w}) \quad (\text{a.s.}, \mu_{\mathbb{R}^d}). \tag{1.23}$$

Although (1.23) is nothing but a reformulation of (1.22), it is an important reformulation. For one thing, if one knows something about the role of *conditioning*, (1.23) has intuitive appeal. To wit, one of the most useful ways to think about the conditional expectation value $\mathbb{E}^{\mathbb{P}}[X|\mathcal{F}'](\omega)$ of X given $\mathcal{F}'$ is as the *best prediction* which one can make about the value of $X(\omega)$ based on whatever information about ω can be gleaned from $\mathcal{F}'$.[20] Thus, (1.23)

[19] If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $X \in L^1(\mathbb{P})$, then the conditional expectation value of X given a σ-algebra $\mathcal{F}' \subseteq \mathcal{F}$ is that $\mathcal{F}'$-measurable element $X' = \mathbb{E}^{\mathbb{P}}[X|\mathcal{F}'] \in L^1(\mathbb{P})$ with the property that $\mathbb{E}^{\mathbb{P}}[X, A] = \mathbb{E}^{\mathbb{P}}[X', A]$ for all $A \in \mathcal{F}'$. When $X \in L^2(\mathbb{P})$, $\mathbb{E}^{\mathbb{P}}[X|\mathcal{F}']$ is simply the orthogonal projection of X onto the subspace of $\mathcal{F}'$-measurable elements of $L^2(\mathbb{P})$. Of course, $\mathbb{E}^{\mathbb{P}}[X|\mathcal{F}']$ is determined only up to a $\mathcal{F}'$-measurable $\mathbb{P}$-null set. See §5.1 in [35] for more details.

[20] When $X \in L^2(\mathbb{P})$, this interpretation gains precision from the fact that $\mathbb{E}^{\mathbb{P}}[X|\mathcal{F}]$ is the $\mathcal{F}'$-measurable element of $L^2(\mathbb{P})$ which is closest, in the sense of $L^2(\mathbb{P})$, to X.

is saying that, based on information about $\mathbf{w}$ available in $\mathcal{B}_t$, the best $\mu_{\mathbb{R}^d}$-prediction which one can make about the value of $M_f^{\Delta_{\mathbb{R}^d}}(t+\tau,\mathbf{w})$ is that it will be $M_f^{\Delta_{\mathbb{R}^d}}(t,\mathbf{w})$. In other words, based on (essentially perfect) information about the path $\mathbf{w} \restriction [0,t]$, the best prediction that one can make about the change in $M_f^{\Delta_{\mathbb{R}^d}}(\,\cdot\,,\mathbf{w})$ between t and $t+\tau$ is that there is none. Alternatively, $t \rightsquigarrow M_f^{\Delta_{\mathbb{R}^d}}(t,\mathbf{w})$ is *conditionally constant* relative of the σ-algebras $t \rightsquigarrow \mathcal{B}_t$.

Random functions which are conditionally constant in the sense just described play a prominent role in the modern theory of probability, where they are called *martingales*. More precisely (cf. §5.2 and §7.1 in [35] for more information), given a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and a non-decreasing family $\{\mathcal{F}_t : t \in [0,\infty)\}$ of sub σ-algebras of $\mathcal{F}$, one says that the triple $\big(X(t),\mathcal{F}_t,\mathbb{P}\big)$ is a martingale on Ω if $X : [0,\infty)\times\Omega \longrightarrow \mathbb{R}$ is a map with the properties that

(1) it is *adapted* to $\{\mathcal{F}_t : t \in [0,\infty)\}$ in the sense that, for each $t \in [0,\infty)$, $X(t)$ is a $\mathcal{B}_t$-measurable element of $L^1(\mathbb{P};\mathbb{R})$,
(2) it is *right-continuous* in the sense that, for each $\omega \in \Omega$, $t \rightsquigarrow X(t,\mathbf{w})$ is right-continuous,
(3) it is conditionally constant in the sense that

$$X(t_1) = \mathbb{E}^{\mathbb{P}}\left[X(t_2)\,\middle|\,\mathcal{F}_{t_1}\right] \quad \mathbb{P}\text{-a.s. for all } 0 \le t_1 < t_2.$$

In situations, like the one here, where the sample space is more or less fixed, like $\mathfrak{W}(\mathbb{R}^d)$, and there is a naturally associated family of sub σ-algebras, like (cf. (1.7)) $\{\mathcal{B}_t : t \in [0,\infty)\}$, we will usually not write out the triple and will simply say something like "$M(t)$ is a $\mathbb{P}$-martingale," when we mean that $M(t)$ is a martingale on $\mathfrak{W}(\mathbb{R}^d)$ relative to $\{\mathcal{B}_t : t \ge 0\}$. Thus, because it entails nothing but a change in the vocabulary which we used in §1.4.1, we can now say that Wiener measure is the unique Borel probability measure ν on $\mathfrak{W}(\mathbb{R}^d)$ with the property that $M_f^{\Delta_{\mathbb{R}^d}}(t)$ is a ν-martingale for every $f \in C_{\mathrm{b}}^{\infty}(\mathbb{R}^d;\mathbb{R})$. In fact, one has the following refinement.

1.24 Theorem. *Wiener measure $\mu_{\mathbb{R}^d}$ is the unique Borel probability measure ν on $\mathfrak{W}(\mathbb{R}^d)$ with the property that $M_f^{\Delta_{\mathbb{R}^d}}(t)$ is a ν-martingale for every $f \in C_{\mathrm{c}}^{\infty}(\mathbb{R}^d;\mathbb{R})$. Moreover, if, for given $f \in C^{1,2}([0,\infty)\times\mathbb{R}^d;\mathbb{R})$,*

$$M_f^{\Delta_{\mathbb{R}^d}}(t,\mathbf{w}) \equiv f\big(t,\mathbf{w}(t)\big) - \int_0^t \left(\frac{\partial f}{\partial \tau} + \tfrac{1}{2}\Delta_{\mathbb{R}^d} f\right)\big(\tau,\mathbf{w}(\tau)\big)\,d\tau,$$

and if, for some $T \in (0,\infty)$ and $p \in (1,\infty)$,

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]}\left|M_f^{\Delta_{\mathbb{R}^d}}(t,\mathbf{w})\right|^p\right] < \infty,$$

then $\big(M_f^{\Delta_{\mathbb{R}^d}}(t\wedge T),\mathcal{B}_t,\mu_{\mathbb{R}^d}\big)$ is a martingale. In particular, this will be the case if

$$\big|f(t,y)\big| + \left|\frac{\partial f}{\partial t} + \tfrac{1}{2}\Delta_{\mathbb{R}^d} f(t,y)\right| \le K\exp\left(\frac{\epsilon|y|^2_{\mathbb{R}^d}}{2T}\right)$$

for all $(t,y) \in [0,T]\times\mathbb{R}^d$ and some $K \in (0,\infty)$ and $\epsilon \in (0,1)$.

PROOF: We begin by checking that if $M_f^{\Delta_{\mathbb{R}^d}}(t)$ is a ν-martingale for every $f \in C_c^\infty(\mathbb{R}^d;\mathbb{R})$, then it is for every $f \in C_c^\infty([0,\infty) \times \mathbb{R}^d;\mathbb{R})$. To this end, let $0 \le t_1 < t_2 < \infty$ and $A \in \mathcal{B}_{t_1}$ be given. Then, since temporal and spacial derivatives commute, one can use (1.22) to justify

$$
\begin{aligned}
&\mathbb{E}^\nu\Big[f\big(t_2,\mathbf{w}(t_2)\big) - f\big(t_1,\mathbf{w}(t_1)\big),\, A\Big] \\
&= \mathbb{E}^\nu\left[\int_{t_1}^{t_2}\left(\frac{\partial f}{\partial \tau}\big(\tau,\mathbf{w}(t_2)\big) + \tfrac{1}{2}\Delta_{\mathbb{R}^d} f\big(t_1,\mathbf{w}(\tau)\big)\right)d\tau,\, A\right] \\
&= \mathbb{E}^\nu\left[\int_{t_1}^{t_2}\left(\frac{\partial f}{\partial \tau}\big(\tau,\mathbf{w}(\tau)\big) + \tfrac{1}{2}\Delta_{\mathbb{R}^d} f\big(\tau,\mathbf{w}(\tau)\big)\right)d\tau,\, A\right] \\
&\qquad + \mathbb{E}^\nu\left[\iint\limits_{t_1\le\tau'\le\tau\le t_2} \tfrac{1}{2}\Delta_{\mathbb{R}^d}\frac{\partial f}{\partial \tau}\big(\tau,\mathbf{w}(\tau')\big)\,d\tau' d\tau,\, A\right] \\
&\qquad\qquad - \mathbb{E}^\nu\left[\iint\limits_{t_1\le\tau'\le\tau\le t_2} \frac{\partial \frac{1}{2}\Delta_{\mathbb{R}^d} f}{\partial \tau}\big(\tau,\mathbf{w}(\tau')\big)\,d\tau' d\tau,\, A\right] \\
&= \mathbb{E}^\nu\left[\int_{t_1}^{t_2}\left(\frac{\partial f}{\partial \tau}\big(\tau,\mathbf{w}(\tau)\big) + \tfrac{1}{2}\Delta_{\mathbb{R}^d} f\big(\tau,\mathbf{w}(\tau)\big)\right)d\tau,\, A\right].
\end{aligned}
$$

After re-arrangement of terms, it becomes clear that the preceding proves that $M_f^{\Delta_{\mathbb{R}^d}}(t)$ is a ν-martingale.

We next note that, by completely standard mollification and truncation procedures, the preceding extends to $f \in C_b^{1,2}([0,\infty) \times \mathbb{R}^d;\mathbb{R})$. In particular, of course, this means that $\nu = \mu_{\mathbb{R}^d}$ as soon as one knows that $M_f^{\Delta_{\mathbb{R}^d}}(t)$ is a ν-martingale for all $f \in C_c^\infty(\mathbb{R}^d;\mathbb{R})$.

To complete the proof when $f \in C^{1,2}([0,\infty) \times \mathbb{R}^d;\mathbb{R})$ and $M_f^{\frac{1}{2}\Delta}$ satisfies the stated integrability property for some T and p, choose a cut-off function $\psi \in C_c^\infty(\mathbb{R};[0,1])$ so that $\psi \restriction [-1,1] = 1$, and, for $R \in (0,\infty)$, set

$$
f_R(t,y) = \psi\left(\frac{t}{T}\right)\psi\left(\frac{|y|_{\mathbb{R}^d}}{R}\right) f(t,y).
$$

Then $M_{f_R}^{\Delta_{\mathbb{R}^d}}(t)$ is a $\mu_{\mathbb{R}^d}$-martingale for every R. Moreover, if

$$
\zeta_R(\mathbf{w}) = \Big(\inf\big\{t \ge 0 : \big|\mathbf{w}(t)\big|_{\mathbb{R}^d} \ge R\big\}\Big) \wedge T,
$$

then $M_f^{\Delta_{\mathbb{R}^d}}(t\wedge\zeta_R) = M_{f_R}^{\Delta_{\mathbb{R}^d}}(t\wedge\zeta_R)$, and so, by Doob's Stopping Time Theorem (cf. Corollary 7.1.15 in [35]), $M_f^{\Delta_{\mathbb{R}^d}}(t \wedge \zeta_R)$ is a $\mu_{\mathbb{R}^d}$-martingale. But, by the integrability hypothesis, $\{M_f^{\Delta_{\mathbb{R}^d}}(t \wedge \zeta_R) : R > 0\}$ is uniformly $\mu_{\mathbb{R}^d}$-integrable, and so the martingale property for $M_f^{\Delta_{\mathbb{R}^d}}(t \wedge T)$ follows after a trivial passage to the limit as $R \nearrow \infty$.

It remains to check the final statement, and for this purpose it suffices to show that, for each $\epsilon \in (0,1)$ and $T \in (0,\infty)$,

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]} \exp\left(\frac{\epsilon|\mathbf{w}(t)|^2_{\mathbb{R}^d}}{2T}\right)\right] \le e(1-\epsilon)^{-\frac{d}{2}}. \tag{1.25}$$

But, for $0 \le t_1 < t_2$,

$$\begin{aligned}\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\exp\Big(\big(\boldsymbol{\theta}, \mathbf{w}(t_2) - \mathbf{w}(t_1)\big)_{\mathbb{R}^d}\Big)\,\Big|\,\mathcal{B}_{t_1}\Big] &= \int_{\mathbb{R}^d} e^{(\boldsymbol{\xi},y)_{\mathbb{R}^d}} g^{(d)}_{t_2-t_1}(y)\,dy \\ &= \exp\left(\frac{(t_2-t_1)|\boldsymbol{\theta}|^2_{\mathbb{R}^d}}{2}\right)\end{aligned}$$

$\mu_{\mathbb{R}^d}$-almost surely. In other words,

$$\exp\left(\big(\boldsymbol{\theta}, \mathbf{w}(t)\big)_{\mathbb{R}^d} - \frac{t|\boldsymbol{\theta}|^2_{\mathbb{R}^d}}{2}\right)$$

is a $\mu_{\mathbb{R}^d}$-martingale for every $\boldsymbol{\theta} \in \mathbb{R}^d$. Hence, by Doob's Inequality (cf. Theorem 7.1.8 in [35]), for any $q \in (1,\infty)$,

$$\begin{aligned}&\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]} \exp\Big(q\big(\boldsymbol{\theta}, \mathbf{w}(t)\big)_{\mathbb{R}^d}\Big)\right] \\ &\quad\le \exp\left(\frac{q|\boldsymbol{\theta}|^2_{\mathbb{R}^d}T}{2}\right)\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]} \exp\left(\big(\boldsymbol{\theta}, \mathbf{w}(t)\big)_{\mathbb{R}^d} - \frac{|\boldsymbol{\theta}|^2_{\mathbb{R}^d}t}{2}\right)^q\right] \\ &\quad\le \left(\frac{q}{q-1}\right)^q \exp\left(\frac{q|\boldsymbol{\theta}|^2_{\mathbb{R}^d}T}{2}\right)\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\exp\left(\big(\boldsymbol{\theta}, \mathbf{w}(T)\big)_{\mathbb{R}^d} - \frac{|\boldsymbol{\theta}|^2_{\mathbb{R}^d}T}{2}\right)^q\right] \\ &\quad= \left(\frac{q}{q-1}\right)^q \exp\left(\frac{q^2|\boldsymbol{\theta}|^2_{\mathbb{R}^d}T}{2}\right).\end{aligned}$$

Thus, if $\tau = \epsilon\big(q^2T\big)^{-1}$ and we integrate both sides of the preceding with respect to $g^{(d)}_\tau(\boldsymbol{\theta})\,d\boldsymbol{\theta}$, then standard Gaussian computations yield

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]} \exp\left(\frac{\epsilon|\mathbf{w}(t)|^2_{\mathbb{R}^d}}{2T}\right)\right] \le \left(\frac{q}{q-1}\right)^q (1-\epsilon)^{-\frac{d}{2}}$$

for any $q \in (1,\infty)$. We get (1.25) from here by letting $q \nearrow \infty$. □

1.4.3. A Heuristic Interpretation. It is sometimes useful and only slightly misleading to interpret the result just stated as saying *Brownian paths are integral curves of* $\frac{1}{2}\Delta_{\mathbb{R}^d}$. To explain, recall that if X is a "nice" (i.e., smooth and reasonably bounded) vector field, then $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$ is the integral curve of X starting at the origin if and only if[21]

$$t \in [0,\infty) \longmapsto f\big(\mathbf{w}(t)\big) - \int_0^t [Xf]\big(\mathbf{w}(\tau)\big)\, d\tau \in \mathbb{R}$$

is (literally) *constant* for every $f \in C_{\mathrm{b}}^{\infty}(\mathbb{R}^d;\mathbb{R})$. Thus, the presence of the adjective *stochastic* is the "only" difference between our conclusion about the relationship of Brownian paths to $\frac{1}{2}\Delta_{\mathbb{R}^d}$ and the preceding characterization of integral curves for a vector field.

Of course, this only difference is a huge difference! Indeed, because it fails to satisfy Leibnitz's rule, $\frac{1}{2}\Delta_{\mathbb{R}^d}$ simply cannot have any integral curves: if $\mathbf{w}$ were one, then we could compute Δf via

$$[\Delta f]\big(\mathbf{w}(0)\big) = 2\lim_{t\searrow 0} \frac{f\big(\mathbf{w}(t)\big) - f\big(\mathbf{w}(0)\big)}{t}$$

and would come to the conclusion that $\big[\Delta_{\mathbb{R}^d} f^2\big](\mathbf{0}) = 2f(\mathbf{0})\big[\Delta_{\mathbb{R}^d} f\big](\mathbf{0})$, which is certainly false in general. That is, it is only because of the flexibility afforded by the notion of "stochastic constancy" that we are able to think of Brownian paths as integral curves of $\frac{1}{2}\Delta_{\mathbb{R}^d}$; and, in view of this, one might be inclined to dismiss this whole idea as utter nonsense. On the other hand, elementary facts (cf. Theorem 7.1.20 and Exercise 7.1.31 in [35]) about martingales allow one to show that, when dealing with operators which do satisfy Leibnitz's rule, there is no difference between the stochastic notion of integral curve and the literal one. Thus (cf. §2.2.1 for a very pedestrian verification), for example, if μ is a probability measure on $\mathfrak{W}(\mathbb{R}^d)$ with the property that

$$f\big(\mathbf{w}(t)\big) - \int_0^t [Xf]\big(\mathbf{w}(\tau)\big)\, d\tau$$

is a μ-martingale for every $f \in C_{\mathrm{b}}^{\infty}(\mathbb{R}^d;\mathbb{R})$, then μ is the point mass concentrated at the one and only integral curve of X starting at $\mathbf{0}$.

Having delved, perhaps too deeply, into the ideas underlying these heuristics, I will conclude with an example which I think demonstrates the power of this line of thinking. Namely, consider a bounded, connected open set G in $\mathbb{R}^d$ which contains $\mathbf{0}$, and suppose that $u \in C(\bar{G};\mathbb{R})$ is harmonic in G. Next, define

$$\zeta^G(\mathbf{w}) = \inf\{t \geq 0 : \mathbf{w}(t) \notin G\} (\equiv \infty \text{ if } \mathbf{w} \subseteq G).$$

Then, in whatever sense Brownian paths are integral curves of $\frac{1}{2}\Delta_{\mathbb{R}^d}$, u should be constant along $\mathbf{w} \restriction \big[0, \zeta^G(\mathbf{w})\big)$. In particular, it is reasonable to hope that

$$u(\mathbf{0}) = \mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[u\Big(\mathbf{w}\big(t \wedge \zeta^G(\mathbf{w})\big)\Big)\Big] \quad \text{for all } t \in [0,\infty).$$

[21] In the following, we associate X with the "directional derivative" operator which it determines.

That this hope is well-founded is just another application of Doob's Stopping Time Theorem. Hence, if one uses the easily verified (cf. Theorem 7.2.11 or Lemma 7.2.18 in [35]) fact that $\zeta^G(\mathbf{w}) < \infty$ $\mu_{\mathbb{R}^d}$-almost surely, then one arrives at the remarkable fact that

$$u(\mathbf{0}) = \mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[f\Big(\mathbf{w}\big(\zeta^G(\mathbf{w})\big)\Big),\ \zeta^G(\mathbf{w}) < \infty\Big],$$

where $f \equiv u \restriction \partial G$ is the boundary value of u. In other words, the distribution of the first place that a Brownian paths exits G is the harmonic measure for G based at the origin. (See Chapter 8 of [35] for more details.)

CHAPTER 2

Diffusions in Euclidean Space

Theorem 1.24 tells us that Wiener measure can be characterized in terms of the operator $\frac{1}{2}\Delta_{\mathbb{R}^d}$. The goal of this chapter is to explain how to construct measures on pathspace which bear the analogous relation to other second order, (possibly degenerate) elliptic differential operators. That is, given $\mathcal{L}$ acting on $C^\infty(\mathbb{R}^N;\mathbb{R})$, we will, under suitable conditions, show how to construct, for each $x \in \mathbb{R}^N$, a probability measure $\mathbb{P}_x^{\mathcal{L}}$ on the Fréchet space $\mathcal{P}(\mathbb{R}^N) \equiv C([0,\infty);\mathbb{R}^N)$ so that $\mathbb{P}_x^{\mathcal{L}}(p(0) = x) = 1$ and, for a sufficiently rich class of $f \in C^\infty(\mathbb{R}^N;\mathbb{R})$,

$$M_f^{\mathcal{L}}(t,p) \equiv f(p(t)) - \int_0^t [\mathcal{L}f](p(\tau))\,d\tau \tag{2.1}$$

is a $\mathbb{P}_x^{\mathcal{L}}$-martingale relative to

$$\mathcal{F}_t \equiv \sigma(\{p(s) : s \in [0,t]\}), \quad t \in [0,\infty). \tag{2.2}$$

Such a measure will be said to *solve the martingale problem for $\mathcal{L}$ starting at x*. The procedure with which we will construct solutions to martingales problems is based on ideas developed in [39].

2.1 Martingale Problems for Operators in Hörmander Form

Although a construction of the sort just described can be and has been (cf. [42]) carried out in considerable generality, we will restrict our attention to the special case when $\mathcal{L}$ is presented in the form[1]

$$\mathcal{L} = X_0 + \frac{1}{2}\sum_{k=1}^{d} X_k^2, \tag{2.3}$$

for some $d \in \mathbb{N}$ and (real) vector fields X_k on $\mathbb{R}^N$. This is the presentation made famous by L. Hörmander in his classic article [21]; and it is important to

[1] Recall that we identify a vector field with the directional derivative operator which it determines. Thus, if X is a vector field, then X^2 is the second order differential operator obtained by two applications of the first order operator X.

realize that, when re-expressed in the classical form with respect to a standard (i.e., orthonormal, linear) coordinate system for $\mathbb{R}^N$, $\mathcal{L}$ is given by

$$\mathcal{L} = \frac{1}{2}\sum_{i,j=1}^{N} a^{i,j}(x)\frac{\partial^2}{\partial\xi^i\partial\xi^j} + \sum_{i=1}^{N} b^i(x)\frac{\partial}{\partial\xi^i}, \tag{2.4}$$

where, for each $x \in \mathbb{R}^N$, the coefficient matrix $a(x)$ is symmetric and (possibly degenerate) non-negative definite. To check these properties of $a(x)$, let $\sigma(x)$ denote the $N \times d$-matrix whose kth column is the vector[2] $\left(X_k^1(x), \dots, X_k^N(x)\right)^\top$ consisting of the coefficients (with respect to the chosen coordinates) of $(X_k)_x$. Then

$$a(x) = \sigma(x)\sigma(x)^\top \quad \text{and } b^i(x) = X_0^i(x) + \frac{1}{2}\sum_{k=1}^{d}(X_k)_x\sigma_k^i, \tag{2.5}$$

which makes the symmetry and non-negativity of $a(x)$ explicit. In particular, *$a(x)$ is non-degenerate precisely when $\{(X_1)_x, \dots, (X_d)_x\}$ spans the tangent space at x.*

As we will see, the advantage gained by writing $\mathcal{L}$ in Hörmander form makes itself apparent as soon as one tries changing coordinates. Namely, in a new coordinate system, the vector fields entering the Hörmander form of $\mathcal{L}$ transform just as they should (i.e., as vector fields do in differential geometry). Hence, the Hörmander form is, from the differential geometric standpoint, *intrinsic*. This fact should be contrasted to the situation in (2.4), where $a(x)$ transforms nicely as a covariant 2-tensor, but the transformation properties of $\mathbf{b}(x)$ are dreadful. The price which one pays for Hörmander form is lack of uniqueness. In fact, if $\mathcal{L}$ is given by (2.4), then for each $d \in \mathbb{N}$, and each choice of smooth $x \rightsquigarrow \sigma(x) \in \mathbb{R}^N \otimes \mathbb{R}^d$ satisfying $a(x) = \sigma(x)\sigma(x)^\top$, one gets, by reversing the preceding procedure, a different Hörmander form expression for $\mathcal{L}$. That is, take $(X_k)_x$, $k = 1, \dots, d$, to be the kth column of $\sigma(x)$ and check that $\mathcal{L} - \frac{1}{2}\sum_{k=1}^d X_k^2$ determines a vector field X_0. Worse, even if one is willing to live with this ambiguity, it turns out that the existence of a smooth $x \rightsquigarrow \sigma(x)$ can pose an insurmountable problem in cases when $a(x)$ is allowed to degenerate.[3] However these (very serious) objections to Hörmander's form will not bother us since, for the most part, the $\mathcal{L}$'s with which we will be dealing admit a "natural" presentation in Hörmander's form.

2.2 The Abelian Case

The strategy on which we will base our construction of solutions to the martingale problem for $\mathcal{L}$ will be to take maximal advantage of the differential geometric invariance properties alluded to in the last paragraph of §2.1.

[2] $A^\top$ is the transpose of the matrix A.

[3] The problem here has its origins in the observation, due to D. Hilbert, that there exist non-negative polynomials which fail to admit *any* representation as a sum of squares of real polynomials.

In particular, we will attempt to the construct solution to the new martingale problems by making a clever mapping from the cases which we already understand, the one for $\frac{1}{2}\Delta_{\mathbb{R}^d}$.

2.2.1. A Single Vector Field. We begin with the probabilistically trivial case when $\mathcal{L}$ is just the first order operator X, where the vector field X is *forward complete* in the sense that, for each $x \in \mathbb{R}^N$, the solution to

$$\frac{dE(\xi,x)}{d\xi} = X_{E(\xi,x)} \quad \text{with } E(0,x) = x \tag{2.6}$$

never *explodes* in the future. That is, for each $x \in \mathbb{R}^N$, the solution to (2.6) exists for all $\xi \in [0,\infty)$. Thus, by the elementary theory of ordinary differential equations, the forward *flow* $\xi \in [0,\infty) \longmapsto E(\xi,\,\cdot\,) \in C^\infty(\mathbb{R}^N;\mathbb{R}^N)$ is a one-parameter semigroup of non-singular, smooth maps of $\mathbb{R}^N$. That is, for all $\xi,\eta \in [0,\infty)$, $E(\xi+\eta,x) = E\big(\eta, E(\xi,x)\big)$.

To see that the martingale problem is trivial in this case, simply observe that, for any continuously differentiable $f : \mathbb{R}^N \longrightarrow \mathbb{R}$,

$$f\big(E(t_2,x)\big) - f\big(E(t_1,x)\big) = \int_{t_1}^{t_2} \big[Xf\big]\big(E(\tau,x)\big)\,d\tau, \quad \text{for all } 0 \le t_1 < t_2.$$

Hence, if we take $\mathbb{P}_x^X$ on $\mathcal{P}(\mathbb{R}^N)$ to be the unit point mass which is concentrated on the path $E(\,\cdot\,,x) \restriction [0,\infty)$, then it is clear that $\mathbb{P}_x^X$ solves the martingale problem for $\mathcal{L}$ starting at x. To see that this is the only solution, suppose that $\mathbb{P}$ is another solution. Given $f \in C_c^\infty(\mathbb{R}^N;\mathbb{R})$, set $u(\xi,y) = f \circ E(\xi,y)$. Then, for any $T > 0$,

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\big[f\big(p(T)\big)\big] - f \circ E(T,x) &= \mathbb{E}^{\mathbb{P}}\Big[u\big(0,p(T)\big) - u\big(T,p(0)\big)\Big] \\ &= \int_0^T \frac{d}{dt}\mathbb{E}^{\mathbb{P}}\Big[u\big(T-t,p(t)\big)\Big]\,dt = 0\end{aligned}$$

since $\frac{d}{d\xi}u(\xi,y) = X_y u(\xi,\,\cdot\,)$ and

$$\frac{d}{dt}\mathbb{E}^{\mathbb{P}}\Big[u\big(\xi,p(t)\big)\Big] = \mathbb{E}^{\mathbb{P}}\Big[X_{p(t)}u(\xi,\,\cdot\,)\Big].$$

Hence, for each $T > 0$, the distribution of $p \in \mathcal{P}(\mathbb{R}^N) \longmapsto p(T) \in \mathbb{R}^N$ under $\mathbb{P}$ is the unit point mass $\delta_{E(T,x)}$. Equivalently, $\mathbb{P}\big(p(T) = E(T,x)\big) = 1$ for every $T \in [0,\infty)$, from which it is an easy step (by countable additivity) first to

$$\mathbb{P}\big(p(t) = E(t,x) \text{ for all rational } t \in [0,\infty)\big) = 1$$

and then (by continuity) to $\mathbb{P}\big(p(\,\cdot\,) = E(\,\cdot\,,x)\big) = 1$. In other words, $\mathbb{P} = \delta_{E(\,\cdot\,,x)}$.

2.2.2. A Single Vector Field Squared. As the preceding makes explicit, the discussion in §1.4.3 is more than mere heuristics when $\frac{1}{2}\Delta_{\mathbb{R}^d}$ is replaced by a vector field. Indeed, for vector fields X, the martingale problem for X is just a stochastic obfuscation of what is an intrinsically deterministic situation.[4] In order to move to a situation in which probability theory plays a more significant role, we next consider the case when $\mathcal{L} = \frac{1}{2}X^2$, where now we take a vector field X which is *complete* in both directions: the solution to (2.6) exists for all $(\xi, x) \in \mathbb{R} \times \mathbb{R}^N$, and therefore $\xi \in \mathbb{R} \longmapsto E(\xi, *) \in C^\infty(\mathbb{R}^N; \mathbb{R}^N)$ is a one-parameter group of diffeomorphisms.

Once again, we try a mapping procedure. Namely, our treatment of the case $\mathcal{L} = X$ relies on the observation that

$$X_{E(\xi,x)}f = \frac{d}{d\xi}\big(f \circ E(\xi, x)\big) = X_x\big(f \circ E(\xi, \cdot\,)\big), \tag{2.7}$$

which led us to map the one-dimensional path $t \rightsquigarrow t$, which corresponds to $\frac{d}{d\xi}$ starting at 0, into the $\mathbb{R}^N$-valued path $t \rightsquigarrow E(t, x)$, which corresponds to X starting from x. This time, we will start from the identity (which follows immediately from (2.7))

$$\frac{1}{2}[X^2 f] \circ E(\xi, x) = \frac{1}{2}\frac{d^2}{d\xi^2}\big(f \circ E(\xi, x)\big),$$

which suggests that we should use $E(\,\cdot\,, x)$ to map the paths corresponding to $\frac{1}{2}\frac{d^2}{d\xi^2}$ in order to get paths which correspond to $\frac{1}{2}X^2$. More precisely, since Brownian paths $w \in \mathfrak{W}(\mathbb{R})$ are the ones associated with $\frac{1}{2}\frac{d^2}{d\xi^2}$, we are guessing that the paths associated with $\frac{1}{2}X^2$ should be given by

$$t \in [0, \infty) \longmapsto p(t, x, w) = E\big(w(t), x\big) \in \mathbb{R}^N.$$

In other words, we are hoping that the solution to the martingale problem for $\frac{1}{2}X^2$ starting from x will be the distribution $\mathbb{P}_x^{\frac{1}{2}X^2}$ of

$$w \in \mathfrak{W}(\mathbb{R}) \longmapsto p(\,\cdot\,, x, w) \in \mathcal{P}(\mathbb{R}) \quad \text{under } \mu_{\mathbb{R}}.$$

Once again, the verification of this *ansatz* is quite easy. First we check that $\mathbb{P}_x^{\frac{1}{2}X^2}$ is a solution. For this purpose, let $s \in [0, \infty)$ and (cf. (1.7)) $A \in \mathcal{B}_s$ be given. We want to show that, for any $f \in C_c^\infty(\mathbb{R}^N; \mathbb{R})$,

$$\begin{aligned}&\mathbb{E}^{\mu_{\mathbb{R}}}\Big[f\big(p(t+s, x, w)\big) - f\big(p(s, x, w)\big),\ A\Big]\\ &\qquad = \mathbb{E}^{\mu_{\mathbb{R}}}\left[\int_0^t \tfrac{1}{2}\big[X^2 f\big]\big(p(\tau + s, x, w)\big)\, d\tau,\ A\right],\end{aligned}$$

[4] Actually, this is slight over-statement. Even for vector fields, the martingale problem formulation can be useful when the vector field is sufficiently bad that it possesses multiple integral curves issuing from the same point.

which, since $\sigma(\{p(\tau,x,w) : \tau \in [0,s]\}) \subseteq \mathcal{B}_s$, is more than enough to prove that $\mathbb{P}_x^{\frac{1}{2}X^2}$ solves the martingale problem. Equivalently, if

$$F(\tau) = \mathbb{E}^{\mu_{\mathbb{R}}}\Big[f\big(p(s+\tau,w)\big),\, A\Big],$$

then we are attempting to prove that

$$F'(\tau) = \mathbb{E}^{\mu_{\mathbb{R}}}\Big[\tfrac{1}{2}\big[X^2 f\big]\big(p(s+\tau,x,w)\big),\, A\Big] \text{ for } \tau \ge 0.$$

Notice that, since $\mathcal{B}_s \subseteq \mathcal{B}_{s+\tau}$, it is sufficient to do this when $\tau = 0$. But, by the group property for the diffeomorphisms $\xi \rightsquigarrow E(\xi, \cdot\,)$,

$$p\big(s+\tau,x,w\big) = E\big([\delta_s w](\tau), p(s,x,w),\big)$$

where $\delta_s : \mathfrak{W}(\mathbb{R}) \longrightarrow \mathfrak{W}(\mathbb{R})$ is defined so that

$$[\delta_s w](\tau) \equiv w(s+\tau) - w(s), \quad \tau \in [0,\infty).$$

Hence, since $w \rightsquigarrow \delta_s w$ is $\mu_{\mathbb{R}}$-independent of $\mathcal{B}_s$ and again has distribution $\mu_{\mathbb{R}}$,

$$F(s+\tau) = \mathbb{E}^{\mu_{\mathbb{R}}}\Big[f\Big(E\big([\delta_s w](\tau), p(s,x,w)\big)\Big),\, A\Big] = \mathbb{E}^{\mu_{\mathbb{R}}}\Big[u\big(\tau, p(s,x,w)\big),\, A\Big],$$

where

$$u(\tau,y) = \frac{1}{\sqrt{2\pi\tau}}\int_{\mathbb{R}} f \circ E(\xi,y)\, e^{-\frac{\xi^2}{2\tau}}\, d\xi.$$

Finally, because

$$\lim_{\tau \searrow 0} \frac{u(\tau,y) - f(y)}{\tau} = \frac{1}{2}\frac{d^2}{d\xi^2} f \circ E(\xi,y)\Big|_{\xi=0} = \tfrac{1}{2}[X^2 f](y),$$

we are done.

To prove that every solution $\mathbb{P}$ must be equal to $\mathbb{P}_x^{\frac{1}{2}X^2}$, again let $f \in C_c^\infty(\mathbb{R}^N;\mathbb{R})$ be given, and, for each $R \in [1,\infty)$, define

$$u_R(t,y) = \psi\left(\frac{|y|}{R}\right)\frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} \psi\left(\frac{\xi}{R}\right) f \circ E(\xi,y)\, e^{-\frac{\xi^2}{2t}}\, d\xi,$$

where $\psi \in C^\infty(\mathbb{R};[0,1])$ is identically 1 on $[-1,1]$ and vanishes off of $(-2,2)$. Using (2.7), one can easily check that u_R is a smooth function on $[0,\infty) \times \mathbb{R}^N$, $u_R(0,y) = \psi(R^{-1}y)f(y)$, and, for some $C < \infty$,

$$\left|\frac{\partial u_R}{\partial t}(t,y) - \frac{1}{2}\big[X^2 u_R(t,\,\cdot\,)\big](y)\right| \le \frac{C}{R}.$$

Hence, if $0 \le t_1 < t_2 < \infty$ and $A \in \mathcal{F}_{t_1}$, then

$$\left|\frac{d}{dt}\mathbb{E}^{\mathbb{P}}\left[u_R\big(t_2 - t, p(t)\big),\, A\right]\right| \le \frac{C}{R} \quad \text{for } t \in (t_1, t_2),$$

and so

$$\left|\mathbb{E}^{\mathbb{P}}\left[\psi\big(R^{-1}p(t_2)\big)f\big(p(t_2)\big),\, A\right] - \mathbb{E}^{\mathbb{P}}\left[u_R\big(t_2 - t_1, p(t_1)\big),\, A\right]\right| \le \frac{C(t_2 - t_1)}{R}.$$

After letting $R \nearrow \infty$, we conclude that

$$\mathbb{E}^{\mathbb{P}}\left[f\big(p(t_2)\big),\, A\right] = \mathbb{E}^{\mathbb{P}}\left[u\big(t_2 - t_1, p(t_1)\big),\, A\right],$$

where u is the function introduced in the preceding paragraph. In other words, we have shown that, for any solution $\mathbb{P}$,

$$\mathbb{E}^{\mathbb{P}}\left[f\big(p(t_2)\big)\,\Big|\,\mathcal{F}_{t_1}\right] = u\big(t_2 - t_1, p(t_1)\big).$$

Starting from the preceding and using induction on $n \ge 1$, one can easily see that, for all $0 < t_1 < \cdots < t_n < \infty$ and $f_1, \dots, f_n \in C_c^\infty(\mathbb{R}^N; \mathbb{R})$,

$$\mathbb{E}^{\mathbb{P}}\left[f_1\big(p(t_1)\big) \cdots f_n\big(p(t_n)\big)\right] = U^{(n)}_{f_1,\dots,f_n}(t_1, \dots, t_n; x),$$

where

$$\begin{aligned} &U^{(n)}_{f_1,\dots,f_n}(t_1, \dots, t_n; x) \\ &= \frac{1}{\sqrt{2\pi t_1}} \int_{\mathbb{R}} f_1\big(E(\xi, x)\big) U^{(n-1)}_{f_2,\dots,f_n}\big(t_2 - t_1, \dots, t_n - t_1; E(\xi, x)\big) e^{-\frac{\xi^2}{2t_1}}\, d\xi \end{aligned}$$

and we have taken $U^{(0)} \equiv 1$. Since this is true of any $\mathbb{P}$ which solves the martingale problem for $\frac{1}{2}X^2$ starting at x, we have now proved that $\mathbb{P}_x^{\frac{1}{2}X^2}$ is the one and only such solution.

2.2.3. Several Commuting Vector Fields. At first, one might hope that the line of reasoning which we have been developing would suffice to handle the general case. More precisely, let $\mathcal{L}$ be given by (2.3), and assume that, for every $\Xi = (\xi^0, \boldsymbol{\xi}) \in [0, \infty) \times \mathbb{R}^d$, the vector field

$$X_\Xi = \xi^0 X_0 + \sum_{k=1}^{d} \xi^k X_k \tag{2.8}$$

is forward complete. Next, define $E(\Xi, x)$ to be the position at time 1 of the integral curve for X_Ξ which passes through x at time 0. Equivalently, $t \in [0, \infty) \longmapsto E(t\Xi, x) \in \mathbb{R}^N$ is determined by

$$\frac{d}{dt}E(t\Xi, x) = \big(X_\Xi\big)_{E(t\Xi, x)} \quad \text{with } E(0\Xi, x) = x. \tag{2.9}$$

In particular, for any $f \in C^2(\mathbb{R}^N;\mathbb{R})$, (cf. (2.3))

$$\left(\frac{\partial}{\partial \xi^0} + \tfrac{1}{2}\Delta_{\mathbb{R}^d}\right) f \circ E(\,\cdot\,,x)\Big|_{(\xi^0,\boldsymbol{\xi})=(0,\mathbf{0})} = [\mathcal{L}f](x), \quad x \in \mathbb{R}^N. \tag{2.10}$$

Unfortunately, in general, (2.10) fails at any $\Xi = (\xi^0, \boldsymbol{\xi})$ other than $(0,\mathbf{0})$. One way to appreciate the difference between what is going on here and what was happening before is to notice that we have lost the semigroup property. Namely, when we were dealing with a single vector field, we could use the semigroup property $E(\xi+\xi',x) = E\big(\xi', E(\xi,x)\big)$ to propagate (2.10). However, the only way to assure that

$$E\big(\Xi + \Xi', x\big) = E\big(\Xi', E(\Xi,x)\big) \tag{2.11}$$

for all Ξ and Ξ' in $[0,\infty) \times \mathbb{R}^d$ is to insist that the vector fields $X_0, \dots, X_d$ commute with one another. Namely, given vector fields Y and Y' with associated (local) flows E and E', the commutator $[Y,Y']$ can be computed at $E(\xi,x)$ by the prescription

$$[Y,Y']_{E(\xi,x)} = \frac{d}{d\xi} E(-\xi,\,\cdot\,)_* Y'_{E(\xi,x)},$$

where, in conformity with the notation from differential geometry, we use $E(\xi,\,\cdot\,)_* Y'_y$ to denote the vector (i.e., directional derivative) at $E(\xi,y)$ which is determined by the formula

$$E(\xi,\,\cdot\,)_* Y'_y f = Y'_y\big(f \circ E(\xi,\,\cdot\,)\big).$$

Hence, if Y and Y' are forward complete, then for any pair $(\xi,\xi') \in [0,\infty)^2$: $[Y,Y'] \equiv \mathbf{0}$ implies $Y'_{E(\xi,y)} = E(\xi,\,\cdot\,)_* Y'_y$ for all $y \in \mathbb{R}^N$, and therefore

$$\begin{aligned}\frac{d}{dt} E\big(t\xi, E'(t\xi',x)\big) &= \xi Y_{E(t\xi,E'(t\xi',x))} + \xi' E(t\xi,\,\cdot\,)_* Y'_{E'(t\xi',x)} \\ &= \xi Y_{E(t\xi,E'(t\xi',x))} + \xi' Y'_{E(t\xi,E'(t\xi',x))} = \big(\xi Y + \xi' Y'\big)_{E(t\xi,E'(t\xi',x))}.\end{aligned}$$

That is,

$$[Y,Y'] \equiv \mathbf{0} \implies t \in [0,\infty) \longmapsto E\big(t\xi, E'(t\xi',x)\big) \in \mathbb{R}^N$$

is the integral curve of $\xi Y + \xi' Y'$ which starts at x. Because the same conclusion follows after reversing the roles of Y and Y', it is evident that $E\big(\xi, E'(\xi',x)\big) = E'\big(\xi', E(\xi,x)\big)$.

In view of the preceding, we see that if E_k denotes the flow generated by X_k, then

$$\begin{aligned}&[X_k, X_\ell] \equiv \mathbf{0} \text{ for } 0 \le k,\ell \le d \\ \implies\ & E\big(\Xi, x\big) = E_{\sigma(0)}\big(\xi^{\sigma(0)}, E_{\sigma(1)}\big(\xi^{\sigma(1)}, \dots, E_{\sigma(d)}\big(\xi^{\sigma(d)}, x\big)\big)\end{aligned}$$

for any permutation σ of $\{0,1,\dots,d\}$ and $\Xi = (\xi^0,\boldsymbol{\xi}) \in [0,\infty) \times \mathbb{R}^d$. In particular,

$$\begin{gathered}[X_k, X_\ell] \equiv \mathbf{0} \text{ for } 0 \le k,\ell \le d \implies \\ \Big[\mathcal{L}\Big(f \circ E(\Xi, *)\Big)\Big](x) = \left(\frac{\partial}{\partial \xi^0} + \tfrac{1}{2}\Delta_{\mathbb{R}^d}\right) f \circ E(\,\cdot\,, x)\Big|_{(\xi^0,\boldsymbol{\xi})=\Xi} = [\mathcal{L}f] \circ E(\Xi, x) \\ \text{for all } \Xi \in [0,\infty) \times \mathbb{R}^d \text{ and } x \in \mathbb{R}^N;\end{gathered}$$

and starting from this, exactly the sort of reasoning used in §§2.2.1 and 2.2.2 leads to the following statement.

2.12 THEOREM. *Let $\{X_k : 0 \le k \le d\}$ be commuting vector fields on $\mathbb{R}^N$, and assume that X_0 is forward complete and that each X_k, $1 \le k \le d$, is complete. Then (cf. (2.8)) X_Ξ is forward complete for every $\Xi \in [0,\infty) \times \mathbb{R}^d$. Moreover, if $\mathcal{L}$ is defined from $\{X_k : 0 \le k \le d\}$ as in (2.3), then, for each $x \in \mathbb{R}^N$, the one and only solution to the martingale problem for $\mathcal{L}$ starting from x is the $\mu_{\mathbb{R}^d}$-distribution[5] $\mathbb{P}_x^{\mathcal{L}}$ of (cf. (2.9))*

$$\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto p(\,\cdot\,, x, \mathbf{w}) \equiv E\Big(\big(\,\cdot\,, \mathbf{w}(\,\cdot\,)\big), x\Big) \in \mathcal{P}(\mathbb{R}^N).$$

2.3 The Non-Abelian Case

Before moving on, the preceding should be understood from the point of view of *controlled* ordinary differential equations. That is, when the the X_k's commute, then another way to describe the path $p(\,\cdot\,, x, \mathbf{w})$ in Theorem 2.12 is as the solution to the ordinary differential equation

$$(2.13)\quad \dot{p}(t,x,\mathbf{w}) = \big(X_0\big)_{p(t,x,\mathbf{w})} + \sum_{k=1}^{d} \dot{w}^k(t)\big(X_k\big)_{p(t,x,\mathbf{w})} \quad \text{with } p(0,x,\mathbf{w}) = x,$$

at least when the *control* $\mathbf{w}$ is smooth. Thus, even when $\mathbf{w}$ is not smooth, $p(\,\cdot\,, x, \mathbf{w})$ can be thought of as a *generalized* solution to (2.13). In terms of this interpretation, the special virtue possessed by commuting vector fields is that this generalization is basically trivial. That is, when the X_k's commute, the map (cf. § 1.2.5) $\mathbf{w} \in H^1(\mathbb{R}^d) \longmapsto p(\,\cdot\,, x, \mathbf{w}) \in \mathcal{P}(\mathbb{R}^N)$, where $p(\,\cdot\,, x, \mathbf{w})$ is the solution to (2.13), admits a unique extention as a continuous map $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto p(\,\cdot\,, x, \mathbf{w}) \in \mathcal{P}(\mathbb{R}^N)$.

When the vector fields do not commute, the situation is radically different and much more complicated. From the point of view adopted in § 2.2, all the additional complications can be traced to the absence of (2.11). From the point of view suggested in the preceding paragraph, the problem is that, although (2.13) can still be interpreted and solved when (cf. § 1.2.5) $\mathbf{w} \in H^1(\mathbb{R}^d)$, the

[5] We have here, and will often below, use the notation ν-distribution to mean the distribution under ν.

map $\mathbf{w} \rightsquigarrow p(\,\cdot\,, x, \mathbf{w})$ is no longer continuous with respect to the topology on $\mathfrak{W}(\mathbb{R}^d)$. Hence, in the non-Abelian case, we no longer have an obvious way to solve (or even interpret) (2.13). Thus, we are forced to develop an approximation scheme and check that, at least for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$, our scheme converges.

2.3.1. The Scheme for Smooth Paths. Until further notice, we will be assuming that we are dealing with vector fields $X_0, \ldots, X_d$ which have bounded derivatives of all orders. That is,[6]

$$\max_{0 \le k \le d} \sup_{x \in \mathbb{R}^N} \left|\partial_x^\alpha X_k\right|_{\mathbb{R}^N} < \infty \quad \text{for all } \alpha \in \mathbb{N}^N. \tag{2.14}$$

In particular, this assumption removes any problems about the long-time existence of integral curves: the elementary theory of ordinary differential equations guarantees that all the vector fields X_Ξ in (2.8) are complete. In fact, (cf. (2.9))

$$(\Xi, x) \in [0, \infty) \times \mathbb{R}^d \times \mathbb{R}^N \longmapsto E(\Xi, x) \in \mathbb{R}^N$$

will be a infinitely differentiable function and, for each $n \in \mathbb{N}$, there exist $C_n \in (0, \infty)$ and $\nu_n \in [0, \infty)$ such that

$$\begin{gathered}\max_{|\alpha| - |\beta| \le n} \left|\partial_\Xi^\alpha \partial_x^\beta \bar{E}(\Xi, x)\right|_{\mathbb{R}^N} \le C_n |\Xi|_{\mathbb{R}^{d+1}}\, e^{\nu_n |\Xi|_{\mathbb{R}^{d+1}}} \\ \text{where } \bar{E}(\Xi, x) \equiv E(\Xi, x) - x.\end{gathered} \tag{2.15}$$

Throughout the discussion which follows, we set $T_{m,n} = m2^{-n}$ for $(m, n) \in \mathbb{N}^2$ and define

$$\begin{aligned}&(t, \mathbf{w}) \in [0, \infty) \times \mathfrak{W}(\mathbb{R}^d) \\ &\qquad \longmapsto \Xi_{(m,n)}(t, \mathbf{w}) = \left(\xi^0_{(m,n)}(t), \boldsymbol{\xi}_{(m,n)}(t, \mathbf{w})\right) \in [0, \infty) \times \mathbb{R}^d\end{aligned}$$

so that

$$\begin{aligned}\xi^0_{(m,n)}(t) &= t \wedge T_{m+1,n} - t \wedge T_{m,n} \\ \boldsymbol{\xi}_{(m,n)}(t, \mathbf{w}) &= \mathbf{w}(t \wedge T_{m+1,n}) - \mathbf{w}(t \wedge T_{m,n})\end{aligned} \qquad \text{for } t \in [T_{m,n}, T_{m+1,n}]. \tag{2.16}$$

With this notation in hand, we can now describe our approximation scheme. Namely, for each $n \in \mathbb{N}$, we define $(t, x, \mathbf{w}) \in [0, \infty) \times \mathbb{R}^N \times \mathfrak{W}(\mathbb{R}^d) \longmapsto p_n(t, x, \mathbf{w}) \in \mathbb{R}^N$ so that

$$\begin{aligned}p_n(0, x, \mathbf{w}) &= x \\ p_n(t, x, \mathbf{w}) &= E\left(\Xi_{(m,n)}(t, \mathbf{w}), p_n(T_{m,n}, x, \mathbf{w})\right)\end{aligned} \tag{2.17}$$

for all $m \in \mathbb{N}$ and $t \in (T_{m,n}, T_{m+1,n}]$. Clearly, for each $n \in \mathbb{N}$ and $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$, $t \rightsquigarrow p_n(t, *, \mathbf{w})$ is continuous as a map from $[0, \infty)$ into $C^\infty(\mathbb{R}^N; \mathbb{R}^N)$. Moreover,

[6] We adopt the notation ∂_x^α to denote the partial derivative $\frac{\partial^{|\alpha|}}{\partial x^\alpha}$ for multi-indices α.

when $\mathbf{w} \in H^1(\mathbb{R}^d)$, it is not all that hard to see that, for each $T \in (0,\infty)$, $\{p_n(\,\cdot\,,*,\mathbf{w}) \restriction [0,T] \times \mathbb{R}^N\}_0^\infty$ converges uniformly to $p(\,\cdot\,,*,\mathbf{w}) \restriction [0,T] \times \mathbb{R}^N$, where $p(t,x,\mathbf{w})$ is the solution to (2.13). To prove this, first note that, by (2.9) and Taylor's Theorem, for any $\varphi \in C^\infty(\mathbb{R}^N;\mathbb{R})$ and $m \in \mathbb{Z}^+$:

$$
\begin{aligned}
\varphi\big(E(\Xi,x)\big) - \varphi(x) &= \sum_{\ell=1}^{m} \frac{1}{\ell!}\big[X_\Xi^\ell \varphi\big](x) \\
&\qquad + \frac{1}{m!}\int_0^1 (1-t)^m \big[X_\Xi^{m+1}\varphi\big]\big(E(t\Xi,x)\big)\,dt
\end{aligned}
\tag{2.18}
$$

In particular, this means (cf. (2.15)) that[7]

$$
\begin{gathered}
\bar{E}(\Xi,x) = E(\Xi,x) - x = \big(X_\Xi\big)_x + R_1(\Xi,x) \\
\text{where } \max_{|\alpha|\le n}\big|\partial_x^\alpha R_1(\Xi,x)\big|_{\mathbb{R}^N} \le C_n |\Xi|^2_{\mathbb{R}^{d+1}} e^{\nu_n|\Xi|_{\mathbb{R}^{d+1}}},
\end{gathered}
\tag{2.19}
$$

and that

$$
\begin{aligned}
&E\big(\Xi', E(\Xi,x)\big) = E(\Xi'+\Xi,x) + \frac{1}{2}\big[X_\Xi, X_{\Xi'}\big]_x + R_2(\Xi,\Xi',x) \\
&\quad\text{where } \max_{|\alpha|\le n}\big|\partial_x^\alpha R_2(\Xi,\Xi',x)\big|_{\mathbb{R}^N} \\
&\qquad\qquad \le C_n\big(|\Xi|^3_{\mathbb{R}^{d+1}} + |\Xi'|^3_{\mathbb{R}^{d+1}}\big)e^{\nu_n(|\Xi|_{\mathbb{R}^{d+1}} + |\Xi'|_{\mathbb{R}^{d+1}})},
\end{aligned}
\tag{2.20}
$$

Now set

$$
\Delta_n(t,x,\mathbf{w}) = p_{n+1}(t,x,\mathbf{w}) - p_n(t,x,\mathbf{w}),
\tag{2.21}
$$

and, for $t \in [T_{m,n}, T_{m+1,n}]$, note that

$$
\begin{aligned}
\Delta_n&(t,x,\mathbf{w}) - \Delta_n(T_{m,n},x,\mathbf{w}) \\
&= \bar{E}\Big(\Xi_{(2m+1,n+1)}(t,\mathbf{w}), E\big(\Xi_{(2m,n+1)}(t,\mathbf{w}), p_{n+1}(T_{m,n},x,\mathbf{w})\big)\Big) \\
&\qquad\qquad - \bar{E}\big(\Xi_{(m,n)}(t,\mathbf{w}), p_n(T_{m,n},x,\mathbf{w})\big) \\
&= E\Big(\Xi_{(2m+1,n+1)}(t,\mathbf{w}), E\big(\Xi_{(2m,n+1)}(t,\mathbf{w}), p_{n+1}(T_{m,n},x,\mathbf{w})\big)\Big) \\
&\qquad\qquad - E\big(\Xi_{(m,n)}(t,\mathbf{w}), p_{n+1}(T_{m,n},x,\mathbf{w})\big) \\
&\qquad + \bar{E}\big(\Xi_{(m,n)}(t,\mathbf{w}), p_{n+1}(T_{m,n},x,\mathbf{w})\big) \\
&\qquad\qquad - \bar{E}\big(\Xi_{(m,n)}(t,\mathbf{w}), p_n(T_{m,n},x,\mathbf{w})\big).
\end{aligned}
$$

[7] The quantities C_n, ν_n, and the like may change from line to line. Thus, there is no implication here, or elsewhere, that they signify more than "the existence of a constant, depending only on n, such that ...".

Hence,

$$
\begin{aligned}
\big|\Delta_n(t,x,\mathbf{w})\big|^2_{\mathbb{R}^N} \le 2\Bigg|\sum_{m<2^nT}\Delta_{(m,n)}(t,x,\mathbf{w})\Bigg|^2_{\mathbb{R}^N} \\
+ 2\Bigg|\sum_{m<2^nT}\tilde{\Delta}_{(m,n)}(x,\mathbf{w})\Bigg|^2_{\mathbb{R}^N},
\end{aligned}
\tag{2.22}
$$

where

$$
\begin{aligned}
\Delta_{(m,n)}(t,x,\mathbf{w}) \equiv \bar{E}\big(\Xi_{(m,n)}(t,\mathbf{w}), p_{n+1}(T_{m,n},x,\mathbf{w})\big) \\
- \bar{E}\big(\Xi_{(m,n)}(t,\mathbf{w}), p_n(T_{m,n},x,\mathbf{w})\big)
\end{aligned}
$$

and

$$
\begin{aligned}
\tilde{\Delta}_{(m,n)}(x,\mathbf{w}) \equiv E\Big(\Xi_{(2m+1,n+1)}(t,\mathbf{w}), E\big(\Xi_{(2m,n+1)}(t,\mathbf{w}), p_{n+1}(T_{m,n},x,\mathbf{w})\big)\Big) \\
- E\big(\Xi_{(m,n)}(t,\mathbf{w}), p_{n+1}(T_{m,n},x,\mathbf{w})\big).
\end{aligned}
$$

Next, we use (2.19) and (2.20) to see that, for $t\in[0,T]$ and $m<2^nT$,

$$
\big|\Delta_{(m,n)}(t,x,\mathbf{w})\big|^2_{\mathbb{R}^N} \le C_T(\mathbf{w})2^{-n}\alpha_{m,n}(\mathbf{w})\big|\Delta_n(T_{m,n},x,\mathbf{w})\big|^2_{\mathbb{R}^N}
$$

and

$$
\big|\tilde{\Delta}_{(m,n)}(x,\mathbf{w})\big|^2_{\mathbb{R}^N} \le C_T(\mathbf{w})4^{-n}\alpha_{m,n}(\mathbf{w})^2,.
$$

where $C_T(\mathbf{w})\in(0,\infty)$ depends only on the constants in (2.19) and (2.20) as well as $\int_0^T(1+|\dot{\mathbf{w}}(t)|^2)\,dt$, and

$$
\alpha_{m,n}(\mathbf{w}) \equiv \int_{T_{m,n}}^{T_{m+1,n}}\Big(1+\big|\dot{\mathbf{w}}(t)\big|^2_{\mathbb{R}^d}\Big)\,dt.
$$

Hence, by Schwarz's inequality,

$$
\Delta_n(T,x,\mathbf{w})^2 \le A_T(\mathbf{w})2^{-n}\sum_{m<2^nT}\Delta_n(T_{m,n},x,\mathbf{w})^2 + B_T(\mathbf{w})4^{-n},
$$

where

$$
\begin{aligned}
A_T(\mathbf{w}) &= 2C_T(\mathbf{w})\int_0^T\Big(1+|\dot{\mathbf{w}}(t)|^2_{\mathbb{R}^d}\Big)\,dt \\
B_T(\mathbf{w}) &= 2C_T(\mathbf{w})\left(\int_0^T\Big(1+|\dot{\mathbf{w}}(t)|^2_{\mathbb{R}^d}\Big)\,dt\right)^2.
\end{aligned}
$$

To complete the proof from here, one only needs the following discrete version of Gronwall's inequality.

2.23 LEMMA. *Let $N \geq 1$ and $\{u_n : 0 \leq n \leq N\} \subseteq [0,\infty)$ be given, and assume that there exist a, $b \in [0,\infty)$ such that*

$$u_n \leq a \sum_{m=0}^{n-1} u_m + b \quad \text{for each } 1 \leq n \leq N.$$

Then $u_N \leq e^{aN}(au_0 + b)$.

PROOF: Set $U_n = \sum_{m=0}^{n} u_m$. Then

$$U_n - U_{n-1} = u_n \leq aU_{n-1} + b, \quad 1 \leq n \leq N,$$

and so, by induction,

$$U_n \leq (1+a)^n u_0 + b \sum_{m=0}^{n-1} (1+a)^m = (1+a)^n u_0 + b\frac{(1+a)^n - 1}{a}$$

and therefore

$$u_N \leq aU_{N-1} + b \leq a(1+a)^{N-1} u_0 + (1+a)^{N-1} b \leq e^{aN}\big(au_0 + b\big). \quad \square$$

Clearly, after combining Lemma 2.23 with the estimates which precede it, we arrive at

$$\sup_{t\in[0,T]} \big|p_{n+1}(t,x,\mathbf{w}) - p_n(t,x,\mathbf{w})\big|_{\mathbb{R}^N} \leq 2^{-\frac{n}{2}} K_T(\mathbf{w}), \quad n \in \mathbb{N},$$

where $K_T(\mathbf{w}) \in (0,\infty)$ depends only on T and $\int_0^T (1+|\dot{\mathbf{w}}(t)|^2)\,dt$. In particular, this means that if $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$ with $\int_0^T |\dot{\mathbf{w}}(t)|^2\,dt < \infty$ for some $T \in (0,\infty)$, then there exists a $p(\,\cdot\,,*,\mathbf{w}) \in C\big([0,T]\times\mathbb{R}^N;\mathbb{R}^N\big)$ to which the sequence $\{p_n(\,\cdot\,,*,\mathbf{w}) \restriction [0,T]\times\mathbb{R}^N : n \geq 0\}$ converges uniformly. Finally, by another application of (2.19), one sees that when (cf. § 1.2.5) $\mathbf{w} \in H^1(\mathbb{R}^d)$,

$$\left| p_n(T,x,\mathbf{w}) - x - \int_0^T \Big((X_0)_{p_n(t,x,\mathbf{w})} + \sum_{k=1}^{d} \dot{w}^k(t)\,(X_k)_{p_n(t,x,\mathbf{w})} \Big)\,dt \right|_{\mathbb{R}^N}$$

tends to 0 as $n \to \infty$. Hence, $p(\,\cdot\,,x,\mathbf{w})$ is indeed the solution to (2.13).

As the preceding makes clear, our scheme works well in the case when $\mathbf{w} \in H^1(\mathbb{R}^d)$. However, although the Feynman picture might encourage one to think otherwise, we saw in §1.2.2 that $\mu_{\mathbb{R}^d}$-almost no $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$ possesses that much smoothness. Thus, our only hope for carrying out this scheme when $\mathbf{w}$ is a generic Brownian path depends on our learning how to remove smoothness from our argument. Of course, smoothness will have to be replaced by something, and, as we are about to see, the replacement is *cancellation*.

2.3.2. The Scheme in the Stochastic Case. [8] The crucial fact which allows us to get away from our reliance of smoothness is *Doob's Inequality* for L^p-martingales. Namely, his inequality says (cf. Theorem 7.1.7 in §7.1 of [35]) that if $(M(t), \mathcal{F}_t, \mathbb{P})$ is a right-continuous, $\mathbb{R}$-valued martingale on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then, for each $p \in (1,\infty)$ and $T \in [0,\infty)$:

$$\mathbb{E}^{\mathbb{P}}\left[\sup_{t\in[0,T]}|M(t)|^p\right]^{\frac{1}{p}} \le \frac{p}{p-1}\mathbb{E}^{\mathbb{P}}\left[|M(T)|^p\right]^{\frac{1}{p}}. \tag{2.24}$$

To see how (2.24) plays a role here, write (cf. the notation in (2.22))

$$\Delta_{(m,n)}(t,x,\mathbf{w}) = \mathbf{M}_{(m,n)}(t,x,\mathbf{w}) + \mathbf{B}_{(m,n)}(t,x,\mathbf{w}) \tag{2.25}$$

where (cf. (2.16))

$$\begin{aligned} &\mathbf{M}_{(m,n)}(t,x,\mathbf{w}) \\ &\equiv \sum_{k=1}^{d} \xi^k_{(m,n)}(t,\mathbf{w})\Big(\big(X_k\big)_{p_{n+1}(T_{m,n},x,\mathbf{w})} - \big(X_k\big)_{p_n(T_{m,n},x,\mathbf{w})}\Big) \end{aligned} \tag{2.26}$$

and (cf. (2.19))

$$\begin{aligned} &\mathbf{B}_{(m,n)}(t,x,\mathbf{w}) \\ &\quad\equiv \xi^0_{(m,n)}(t)\Big(\big(X_0\big)_{p_{n+1}(T_{m,n},x,\mathbf{w})} - \big(X_0\big)_{p_n(T_{m,n},x,\mathbf{w})}\Big) \\ &\qquad + \Big(R_1\big(\Xi_{(m,n)}(t,\mathbf{w}), p_{n+1}(T_{m,n},x,\mathbf{w})\big) \\ &\qquad\qquad - R_1\big(\Xi_{(m,n)}(t,\mathbf{w}), p_n(T_{m,n},x,\mathbf{w})\big)\Big). \end{aligned} \tag{2.27}$$

Next observe that, for each (m,n), $(\mathbf{M}_{(m,n)}(t,x), \mathcal{B}_t, \mu_{\mathbb{R}^d})$ is an $\mathbb{R}^N$-valued martingale. To see this, first note that (cf. Theorem 1.24) $\boldsymbol{\xi}_{(m,n)}(t)$ is an $\mathbb{R}^d$-valued $\mu_{\mathbb{R}^d}$-martingale. Thus, the martingale property for $\mathbf{M}_{(m,n)}(\,\cdot\,,x)$ follows easily from the facts that $\boldsymbol{\xi}_{(m,n)}(t,\mathbf{w}) = \mathbf{0}$ when $t \le T_{m,n}$ and that both the maps $\mathbf{w}\rightsquigarrow p_{n+1}(T_{m,n},x,\mathbf{w})$ and $\mathbf{w}\rightsquigarrow p_n(T_{m,n},x,\mathbf{w})$ are (cf. (1.7)) $\mathcal{B}_{T_{m,n}}$-measurable. As a consequence of these observations, we now see that

$$\mathbf{M}_n(t,x,\mathbf{w}) \equiv \sum_{m=0}^{\infty} \mathbf{M}_{(m,n)}(t,x,\mathbf{w}) \tag{2.28}$$

is also an $\mathbb{R}^d$-valued $\mu_{\mathbb{R}^d}$-martingale. Thus, by[9] (2.24),

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]}\left|\mathbf{M}_n(t,x,\mathbf{w})\right|^2_{\mathbb{R}^N}\right] \le 4\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\left|\sum_{m<2^nT}\mathbf{M}_{(m,n)}(T,x,\mathbf{w})\right|^2_{\mathbb{R}^N}\right].$$

---

[8] The reader who prefers to avoid gory details might want to skip directly to the statement in Theorem 2.40 below.

[9] Notice that (2.24) applies, *mutatis mutandis*, to any martingale with values in a separable Hilbert space.

At the same time, for each $T \in [0, \infty)$,

$$m \in \mathbb{N} \longmapsto \sum_{\ell=0}^{m} \mathbf{M}_{(\ell,n)}(T, x, \mathbf{w}) \in \mathbb{R}^N$$

is a discrete parameter $\mu_{\mathbb{R}^d}$-martingale relative to $\{\mathcal{B}_{T \wedge T_{m,n}} : m \in \mathbb{N}\}$. In particular, this means that

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\left|\sum_{m<2^nT} \mathbf{M}_{(m,n)}(T, x, \mathbf{w})\right|_{\mathbb{R}^N}^2\right] = \sum_{m<2^nT} \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\left|\mathbf{M}_{(m,n)}(T, \mathbf{w})\right|_{\mathbb{R}^N}^2\right].$$

Hence, we now know that

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]}\left|\mathbf{M}_n(t, x, \mathbf{w})\right|_{\mathbb{R}^N}^2\right] \le 4 \sum_{m<2^nT} \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\left|\mathbf{M}_{(m,n)}(T, \mathbf{w})\right|_{\mathbb{R}^N}^2\right].$$

Finally, note that, for some $C < \infty$ depending only on the first derivatives of the X_k's,

$$\begin{aligned}
&\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\left|\mathbf{M}_{(m,n)}(T, x, \mathbf{w})\right|_{\mathbb{R}^N}^2\right] \\
&\quad = \xi^0_{(m,n)}(T) \sum_{k=1}^{d} \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\left|\left(X_k\right)_{p_{n+1}(T_{m,n},x,\mathbf{w})} - \left(X_k\right)_{p_n(T_{m,n},x,\mathbf{w})}\right|_{\mathbb{R}^N}^2\right] \\
&\quad \le C\xi^0_{(m,n)}(T)\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\left|\Delta_n(T_{m,n}, \mathbf{w})\right|_{\mathbb{R}^d}^2\right],
\end{aligned}$$

where we have used the easily verified fact that

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\xi^k_{(m,n)}(t, \mathbf{w})\xi^\ell_{(m,n)}(t, \mathbf{w}) \,\middle|\, \mathcal{B}_{T_{m,n}}\right] = \delta_{k,\ell}\xi^0_{(m,n)}(t) \quad (\text{a.s.}, \mu_{\mathbb{R}^d}). \tag{2.29}$$

After combining these, we obtain

$$\begin{aligned}
&\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]}\left|\mathbf{M}_n(t, x, \mathbf{w})\right|_{\mathbb{R}^N}^2\right] \\
&\qquad \le 4C \sum_{m<2^nT} \xi^0_{(m,n)}(T)\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\left|\Delta_n(T_{m,n}, \mathbf{w})\right|_{\mathbb{R}^d}^2\right].
\end{aligned} \tag{2.30}$$

We now want to develop a similar estimate for (cf. (2.27))

$$\mathbf{B}_n(t, x, \mathbf{w}) \equiv \sum_{m=1}^{\infty} \mathbf{B}_{(m,n)}(t, x, \mathbf{w}). \tag{2.31}$$

But by (2.19), there is a $C < \infty$ such that

$$\left|\mathbf{B}_{(m,n)}(t,x,\mathbf{w})\right|_{\mathbb{R}^N} \le C\Big(\xi^0_{(m,n)}(t) + \left|\boldsymbol{\xi}_{(m,n)}(t,\mathbf{w})\right|^2_{\mathbb{R}^d}\Big) \\ \times \exp\Big(C\left|\Xi_{(m,n)}(t,\mathbf{w})\right|_{\mathbb{R}^{d+1}}\Big)\left|\Delta_n(T_{m,n},x,\mathbf{w})\right|_{\mathbb{R}^N}.$$

Thus, since $\Xi_{(m,n)}(\,\cdot\,,\mathbf{w})$ is $\mu_{\mathbb{R}^d}$-independent of $\Delta_n(T_{m,n},x,\mathbf{w})$, (1.25) combined with Wiener scaling invariance (cf. §1.2.2) lead to the existence of a $C' < \infty$ for which

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]}\left|\mathbf{B}_{(m,n)}(t,x,\mathbf{w})\right|^2_{\mathbb{R}^N}\right] \le C'\xi^0_{(m,n)}(T)^2\,\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\left|\Delta_n(T_{m,n},x,\mathbf{w})\right|^2_{\mathbb{R}^N}\right],$$

and so, by Schwarz's inequality, we obtain

$$\tag{2.32} \begin{aligned} &\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]}\left|\mathbf{B}_n(t,x,\mathbf{w})\right|^2_{\mathbb{R}^N}\right] \\ &\qquad \le CT\sum_{m<2^nT}\xi^0_{(m,n)}(T)\,\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\left|\Delta_n(T_{m,n},x,\mathbf{w})\right|^2_{\mathbb{R}^N}\right] \end{aligned}$$

for yet another choice of $C < \infty$.

To finish the program, we still have to handle the term in (2.22) involving the $\tilde{\Delta}_{(m,n)}$'s, and again we need to begin by splitting off a martingale part. Namely, we write

$$\tag{2.33} \tilde{\Delta}_{(m,n)}(t,x,\mathbf{w}) = \tilde{\mathbf{M}}_{(m,n)}(t,x,\mathbf{w}) + \tilde{\mathbf{B}}_{(m,n)}(t,x,\mathbf{w}),$$

where (cf. (2.20))

$$\tag{2.34} \begin{aligned} &\tilde{\mathbf{M}}_{(m,n)}(t,x,\mathbf{w}) \\ &\equiv \frac{1}{2}\sum_{1\le k\ne \ell\le d}\xi^k_{(2m+1,n+1)}(t,\mathbf{w})\xi^\ell_{(2m,n+1)}(t,\mathbf{w})\left[X_k,X_\ell\right]_{p_{n+1}(T_{m,n},x,\mathbf{w})} \end{aligned}$$

and

$$\tag{2.35} \begin{aligned} &\tilde{\mathbf{B}}_{(m,n)}(t,x,\mathbf{w}) \\ &\equiv \frac{1}{2}\sum_{k=1}^d\Big(\xi^k_{(2m+1,n+1)}(t,\mathbf{w})\xi^0_{(2m,n+1)}(t) \\ &\qquad\qquad - \xi^0_{(2m+1,n+1)}(t)\xi^k_{(2m,n+1)}(t,\mathbf{w})\Big)\left[X_k,X_0\right]_{p_{n+1}(T_{m,n},x,\mathbf{w})} \\ &\quad + R_2\big(\Xi_{(2m,n+1)}(t,\mathbf{w}),\Xi_{(2m+1,n+1)}(t,\mathbf{w}),p_{n+1}(T_{m,n},x,\mathbf{w})\big). \end{aligned}$$

The crucial observation to be made here is that, because they are $\mu_{\mathbb{R}^d}$-independent, the product of $\xi^k_{(2m+1,n+1)}(t,\mathbf{w})$ and $\xi^\ell_{(2m,n+1)}(t,\mathbf{w})$ is again a $\mu_{\mathbb{R}^d}$-martingale. Hence, by exactly the same reasoning as we used to obtain (2.30) and (2.32), we can find a $C<\infty$ for which

$$\text{(2.36)} \qquad \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]}\left|\bar{\mathbf{M}}_n(t,x,\mathbf{w})\right|^2_{\mathbb{R}^N}\right] \le C\sum_{m<2^nT}\xi^0_{(m,n)}(T)^2$$

and

$$\text{(2.37)} \qquad \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]}\left|\bar{\mathbf{B}}_n(t,x,\mathbf{w})\right|^2_{\mathbb{R}^N}\right] \le CT\sum_{m<2^nT}\xi^0_{(m,n)}(T)^2,$$

where

$$\bar{\mathbf{M}}_n(t,x,\mathbf{w}) \equiv \sum_{m=0}^{\infty}\bar{\mathbf{M}}_{(m,n)}(t,x,\mathbf{w})$$

and

$$\bar{\mathbf{B}}_n(t,x,\mathbf{w}) \equiv \sum_{m=0}^{\infty}\tilde{\mathbf{B}}_{(m,n)}(t,x,\mathbf{w}).$$

Since (cf. (2.22))

$$\begin{aligned}\left|\Delta_n(t,x,\mathbf{w})\right|^2_{\mathbb{R}^N} \le 4\left|\mathbf{M}_n(t,x,\mathbf{w})\right|^2_{\mathbb{R}^N} &+ 4\left|\mathbf{B}_n(t,x,\mathbf{w})\right|^2_{\mathbb{R}^N}\\ &+ 4\left|\tilde{\mathbf{M}}_n(t,x,\mathbf{w})\right|^2_{\mathbb{R}^N} + 4\left|\tilde{\mathbf{B}}_n(t,x,\mathbf{w})\right|^2_{\mathbb{R}^N},\end{aligned}$$

(2.30), (2.32), (2.36), and (2.37) combine to give

$$\begin{aligned}&\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]}\left|\Delta_n(t,x,\mathbf{w})\right|^2_{\mathbb{R}^N}\right]\\ &\qquad\le C(1+T)(T\wedge 2^{-n})\left(1+\sum_{m<2^nT}\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\left|\Delta_n(T_{m,n},x,\mathbf{w})\right|^2_{\mathbb{R}^N}\right]\right)\end{aligned}$$

for a choice of $C<\infty$ which depends only on the estimates in (2.19) and (2.20). Thus, after an application of Lemma 2.23, we conclude first that

$$\text{(2.38)} \qquad \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]}\left|\Delta_n(t,x,\mathbf{w})\right|^2_{\mathbb{R}^N}\right] \le Ce^{CT}(T\wedge 2^{-n}),$$

with a $C<\infty$ having the same dependence. Second (via the triangle inequality), we see that, for each $M\in\mathbb{N}$,

$$\text{(2.39)} \quad \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{n\ge M}\sup_{t\in[0,T]}\left|p_n(t,x,\mathbf{w})-p_M(t,x,\mathbf{w})\right|^2_{\mathbb{R}^N}\right] \le 9Ce^{CT}(T\wedge 2^{-M})$$

Hence, elementary measure theoretic considerations bring us to our goal:

2.40 THEOREM. *For each $x \in \mathbb{R}^N$ there exists a Borel measurable map $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto p(\cdot\,, x, \mathbf{w}) \in \mathcal{P}(\mathbb{R}^N)$ such that (cf. (2.17)), for each $T \in (0,\infty)$,*

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \left|p_n(t,x,\mathbf{w}) - p(t,x,\mathbf{w})\right|_{\mathbb{R}^N} = 0$$

both $\mu_{\mathbb{R}^d}$-almost surely and in $L^2(\mu_{\mathbb{R}^d};\mathbb{R})$. In particular, for each $T \in [0,\infty)$,

$$\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto p(\cdot\,, x, \mathbf{w}) \restriction [0,T] \in C\big([0,T];\mathbb{R}^N\big)$$

is measurable with respect to the $\mu_{\mathbb{R}^d}$-completion $\tilde{\mathcal{B}}_T$ of $\mathcal{B}_T$. Finally (cf. (2.3)), for each $T > 0$ and $f \in C^{1,2}_{\mathrm{b}}([0,T]\times\mathbb{R}^N;\mathbb{R})$,

$$f\big(t\wedge T, p(t\wedge T, x, \mathbf{w})\big) - \int_0^{t\wedge T} \left(\frac{\partial f}{\partial\tau} + \mathcal{L}f\right)\big(\tau, p(\tau,x,\mathbf{w})\big)\,d\tau$$

is a $\mu_{\mathbb{R}^d}$-martingale relative to $\{\tilde{\mathcal{B}}_t : t \geq 0\}$. Thus, the $\mu_{\mathbb{R}^d}$-distribution of $\mathbf{w} \rightsquigarrow p(\cdot\,,x,\mathbf{w})$ solves the martingale problem for $\mathcal{L}$ starting from x. (See Theorem 2.68 below for the corresponding uniqueness assertion.)

PROOF: Only the final part of the statement requires further comment. Further, by an easy approximation argument, we may and will assume that $f \in C^\infty_{\mathrm{b}}([0,T]\times\mathbb{R}^N;\mathbb{R})$. Then, by (2.18), for any $T_{m,n} \leq t \leq T_{m+1,n}\wedge T$:

$$\begin{aligned}
&f\big(t, p_n(t,x,\mathbf{w})\big) - f\big(T_{m,n}, p_n(T_{m,n},x,\mathbf{w})\big) - \int_{T_{m,n}}^t \frac{\partial f}{\partial\tau}\big(\tau, p_n(T_{m,n},x,\mathbf{w})\big)\,d\tau \\
&\quad = \big[X_{\boldsymbol{\Xi}_{(m,n)}(t,\mathbf{w})} f(t,\,\cdot\,)\big]\big(p_n(T_{m,n},x,\mathbf{w})\big) \\
&\qquad\qquad + \frac{1}{2}\big[X^2_{(0,\boldsymbol{\xi}_{(m,n)}(t,\mathbf{w}))} f(t,\,\cdot\,)\big]\big(p_n(T_{m,n},x,\mathbf{w})\big) + R_{(m,n)}(t,x,\mathbf{w}),
\end{aligned}$$

where

$$\begin{aligned}
&\big|R_{(m,n)}(t,x,\mathbf{w})\big| \\
&\qquad \leq C\|f\|_{C^3_{\mathrm{b}}([0,T]\times\mathbb{R}^N;\mathbb{R})}\Big(\xi^0_{(m,n)}(t)^2 + \big|\boldsymbol{\xi}_{(m,n)}(t,\mathbf{w})\big|^3_{\mathbb{R}^d}\Big)e^{C|\boldsymbol{\Xi}_{(m,n)}(t,\mathbf{w})|_{\mathbb{R}^{d+1}}}
\end{aligned}$$

for some $C < \infty$. Thus, by (2.29) and (1.25), there is a $C' < \infty$ for which

$$\begin{aligned}
&\bigg|\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[f\big(t,p_n(t,x,\mathbf{w})\big) - f\big(T_{m,n}, p_n(T_{m,n},x,\mathbf{w})\big)\,\Big|\,\tilde{\mathcal{B}}_{T_{m,n}}\Big] \\
&\quad - \int_{T_{m,n}}^t \frac{\partial f}{\partial\tau}\big(\tau, p_n(T_{m,n},x,\mathbf{w})\big)\,d\tau + \xi^0_{(m,n)}(t)\big[\mathcal{L}f(t,\,\cdot\,)\big]\big(p_n(T_{m,n},x,\mathbf{w})\big)\bigg|
\end{aligned}$$

is dominated by $C'2^{-\frac{3n}{2}}$. Now suppose that $t_1 = T_{\ell_1,N} < t_2 = T_{\ell_2,N} \leq T$. Then, for any $A \in \tilde{\mathcal{B}}_{t_1}$ and $n \geq N$:

$$\begin{aligned}
&\bigg|\mathbb{E}^{\mu_{\mathbb{R}^d}}\bigg[f\big(t_2, p_n(t_2,x,\mathbf{w})\big) - f\big(t_1,p_n(t_1,x,\mathbf{w})\big) \\
&\qquad - \sum_{m=2^{n-N}\ell_1}^{2^{n-N}\ell_2-1}\bigg(\int_{T_{m,n}}^{T_{m+1,n}} \frac{\partial f}{\partial\tau}\big(\tau,p_n(t,x,\mathbf{w})\big)\,d\tau \\
&\qquad\qquad + 2^{-n}\big[\mathcal{L}f(T_{m,n},\,\cdot\,)\big]\big(p_n(T_{m,n},x,\mathbf{w})\big)\bigg),\,A\bigg]\bigg| \leq C''T2^{-\frac{n}{2}}.
\end{aligned}$$

Hence, after passing to the limit as $n \to \infty$, we conclude that

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\Bigg[f\big(t_2, p(t_2,x,\mathbf{w})\big) - f\big(t_1, p(t_1,x,\mathbf{w})\big) \\ - \int_{t_1}^{t_2}\left(\frac{\partial f}{\partial \tau} + \mathcal{L}f\right)\big(\tau, p(\tau,x,\mathbf{w})\big)\,d\tau \,\bigg|\, \bar{\mathcal{B}}_{t_1}\Bigg] = 0,$$

first for dyadic and then for all $0 \le t_1 < t_2 \le T$. $\square$

2.3.3. Basic Size Estimates. It is important to know that, in some qualitative sense, the paths $p(\cdot, x, \mathbf{w})$ behave somewhat like the original Wiener paths $\mathbf{w}$. In particular, we will show here that they satisfy estimates which are analogous to (1.25). The proof will rely on the following general fact about the relationship between martingales and processes of bounded variation.[10]

2.41 LEMMA. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}_t : t \ge 0\}$ a non-decreasing family of sub-σ-algebras of $\mathcal{F}$. Further, suppose that X and Y are a pair of $\{\mathcal{F}_t : t \ge 0\}$-adapted[11] functions on $[0,\infty) \times \Omega$ such that $X(\cdot,\omega) \in C([0,\infty);\mathbb{R})$ and $Y(\cdot,\omega) \in C^1([0,\infty);\mathbb{R})$ for every $\omega \in \Omega$. Finally, assume that $(X(t), \mathcal{F}_t, \mathbb{P})$ is a martingale, $Y(0)$ is bounded, and that, for each $T \in (0,\infty)$, $\dot{Y} \restriction [0,T] \times \Omega$ is bounded. Then*

$$\left(X(t)Y(t) - \int_0^t X(\tau)\dot{Y}(\tau)\,d\tau, \mathcal{F}_t, \mathbb{P}\right)$$

is a martingale.

PROOF: Let $0 \le s < t < \infty$ be given, and set $\tau_{m,n} = s + \frac{m}{n}(t-s)$ for $n \ge 1$ and $0 \le m \le n$. Given an $A \in \mathcal{F}_s$, we have

$$\begin{aligned}
&\mathbb{E}\big[X(t)Y(t),\, A\big] - \mathbb{E}\big[X(s)Y(s),\, A\big] \\
&= \sum_{m=1}^{n} \mathbb{E}\big[X(\tau_{m,n})Y(\tau_{m,n}) - X(\tau_{m-1,n})Y(\tau_{m-1,n}),\, A\big] \\
&= \sum_{m=1}^{n} \mathbb{E}\big[X(\tau_{m,n})Y(\tau_{m,n}) - X(\tau_{m,n})Y(\tau_{m-1,n}),\, A\big] \\
&\longrightarrow \mathbb{E}\left[\int_s^t X(\tau)\dot{Y}(\tau)\,d\tau,\, A\right]
\end{aligned}$$

as $n \to \infty$. $\square$

---

[10] See Theorem 7.1.19 and Exercise 7.1.30 in [35] for more on this subject.

[11] A function F on $[0,\infty) \times \Omega$ is said to be $\{\mathcal{F}_t : t \ge 0\}$-adapted if it is $\mathcal{B}_{[0,\infty)} \times \bigvee_{t \ge 0} \mathcal{F}_t$-measurable and, for each $t \in [0,\infty)$, $\omega \rightsquigarrow F(t,\omega)$ is $\mathcal{F}_t$-measurable.

2.42 THEOREM. *Define* $x \in \mathbb{R}^N \longmapsto a(x) = \big((a^{i,j}(x))\big)$ *and* $x \in \mathbb{R}^N \longmapsto \mathbf{b}(x) = \big(b^1(x), \dots, b^N(x)\big) \in \mathbb{R}^N$, *where the* $a^{i,j}$*'s and* b^i*'s are the ones in* (2.4). *Then, for each* $\boldsymbol{\theta} \in \mathbb{R}^N$,

$$M^{\boldsymbol{\theta}}(t,x,\mathbf{w}) \equiv \exp\Bigg[\bigg(\boldsymbol{\theta}, p(t,x,\mathbf{w}) - x - \int_0^t \mathbf{b}\big(p(\tau,x,\mathbf{w})\big)\,d\tau\bigg)_{\mathbb{R}^N} - \frac{1}{2}\int_0^t \Big(\boldsymbol{\theta}, a\big(p(\tau,x,\mathbf{w})\big)\boldsymbol{\theta}\Big)_{\mathbb{R}^N}\,d\tau\Bigg]$$

is a $\mu_{\mathbb{R}^d}$*-martingale relative to* $\{\bar{\mathcal{B}}_t : t \geq 0\}$. *In particular, if we take* $A = \sup_{x\in\mathbb{R}^N} \|a(x)\|_{\mathrm{op}}$, *then, for each* $\epsilon \in (0,1)$ *and all* $(T,x) \in (0,\infty) \times \mathbb{R}^N$,

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]} \exp\left(\frac{\epsilon\left|p(t,x,\mathbf{w}) - x - \int_0^t \mathbf{b}\big(p(\tau,x,\mathbf{w})\big)\,d\tau\right|_{\mathbb{R}^N}^2}{2AT}\right)\right] \leq e(1-\epsilon)^{-\frac{N}{2}}. \tag{2.43}$$

Thus, for each $n \geq 1$,

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]}\left|p(t,x,\mathbf{w}) - x - \int_0^t \mathbf{b}\big(p(\tau,x,\mathbf{w})\,d\tau\right|_{\mathbb{R}^N}^{2n}\right] \leq n!2^{\frac{N}{2}}e(4AT)^n. \tag{2.44}$$

PROOF: Without loss in generality, we take $x = \mathbf{0}$. Next, set

$$p(t,\mathbf{w}) = p(t,\mathbf{0},\mathbf{w}) \quad \text{and} \quad \tilde{p}(t,\mathbf{w}) = p(t,\mathbf{w}) - \int_0^t \mathbf{b}\big(p(\tau,\mathbf{w})\big)\,d\tau.$$

Finally, for $n \geq 1$, set $M_n^{\boldsymbol{\theta}}(t,\mathbf{w}) = M^{\boldsymbol{\theta}}\big(t \wedge \zeta_n(\mathbf{w}),\mathbf{w}\big)$, where $\zeta_n(\mathbf{w}) \equiv \inf\{t \geq 0 : |p(t,\mathbf{w})|_{\mathbb{R}^N} \geq n\}$.

We begin by showing that $M_n^{\boldsymbol{\theta}}(t,\mathbf{w})$ is a $\mu_{\mathbb{R}^d}$-martingale. For this purpose, take $X(t,\mathbf{w})$ equal to

$$\exp\Big(\big(\boldsymbol{\theta}, p(t\wedge\zeta_n(\mathbf{w}),\mathbf{w})\big)_{\mathbb{R}^N}\Big) - \int_0^{t\wedge\zeta_n(\mathbf{w})}\bigg(\Big(\boldsymbol{\theta},\mathbf{b}\big(p(\tau,\mathbf{w})\big)\Big)_{\mathbb{R}^N} + \frac{1}{2}\Big(\boldsymbol{\theta}, a\big(p(\tau,\mathbf{w})\big)\boldsymbol{\theta}\Big)_{\mathbb{R}^N}\bigg) \times \exp\Big(\big(\boldsymbol{\theta}, p(\tau,\mathbf{w})\big)_{\mathbb{R}^N}\Big)\,d\tau$$

and $Y(t,\mathbf{w})$ equal to

$$\exp\left(-\int_0^{t\wedge\zeta_n(\mathbf{w})}\bigg(\Big(\boldsymbol{\theta},\mathbf{b}\big(p(\tau,\mathbf{w})\big)\Big)_{\mathbb{R}^N} + \frac{1}{2}\Big(\boldsymbol{\theta}, a\big(p(\tau,\mathbf{w})\big)\boldsymbol{\theta}\Big)_{\mathbb{R}^N}\bigg)\,d\tau\right).$$

By Theorem 2.40 (together with the sort of cut-off procedure used in its proof) applied to the function $x \rightsquigarrow \exp\big((\boldsymbol{\theta}, x)_{\mathbb{R}^N}\big)$, we know that $X(t, \mathbf{w})$ is a $\mu_{\mathbb{R}^d}$-martingale. Moreover, a little integration by parts shows that

$$M_n^{\boldsymbol{\theta}}(t, \mathbf{w}) = X(t, \mathbf{w})Y(t, \mathbf{w}) - \int_0^t X(\tau, \mathbf{w})\dot{Y}(\tau, \mathbf{w})\, d\tau.$$

Hence, by Lemma 2.41, we know that $M_n^{\boldsymbol{\theta}}(t, \mathbf{w})$ is also a $\mu_{\mathbb{R}^d}$-martingale.

We next apply Doob's Inequality (cf. (2.24)) to see that, for any $q \in (1, \infty)$,

$$\begin{aligned}
&\exp\left(-\frac{qA|\boldsymbol{\theta}|^2_{\mathbb{R}^N} T}{2}\right) \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T\wedge\zeta_n(\mathbf{w})]} \exp\Big(q\big(\boldsymbol{\theta}, \tilde{p}(t, \mathbf{w})\big)_{\mathbb{R}^N}\Big)\right] \\
&\quad \le \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]} M_n^{\boldsymbol{\theta}}(t, \mathbf{w})^q\right] \le \left(\frac{q}{q-1}\right)^q \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[M_n^{\boldsymbol{\theta}}(T, \mathbf{w})^q\right] \\
&\quad \le \left(\frac{q}{q-1}\right)^q \exp\left(\frac{q(q-1)A|\boldsymbol{\theta}|^2_{\mathbb{R}^N} T}{2}\right) \mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[M_n^{q\boldsymbol{\theta}}(T, \mathbf{w})\Big].
\end{aligned}$$

Because $M_n^{q\boldsymbol{\theta}}(t, \mathbf{w})$ is a martingale with mean value 1, the fact that the preceding holds for all $n \ge 1$ allows us to conclude that

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]} \exp\Big(q\big(\boldsymbol{\theta}, \tilde{p}(t, \mathbf{w})\big)_{\mathbb{R}^N}\Big)\right] \le \left(\frac{q}{q-1}\right)^q \exp\left(\frac{q^2 A|\boldsymbol{\theta}|^2_{\mathbb{R}^N} T}{2}\right)$$

after letting $n \to \infty$ and applying Fatou's Lemma. Starting from the preceding and using the same argument as we used in the last step in the derivation of (1.25), we arrive at (2.43).

Finally, knowing (2.43), we have more than enough integrability to see that $M_n^{\boldsymbol{\theta}}(t, \mathbf{w}) \longrightarrow M^{\boldsymbol{\theta}}(t, \mathbf{w})$ in $L^1(\mu_{\mathbb{R}^d}; \mathbb{R})$. Hence, the martingale property for $M^{\boldsymbol{\theta}}(t, \mathbf{w})$ follows from that for the $M_n^{\boldsymbol{\theta}}(t, \mathbf{w})$'s. □

Given (2.43), exactly the same argument as we used in the proof of analogous fact in Theorem 1.24 can now be applied to prove the next statement.

2.45 COROLLARY. *Assume that* $f \in C^{1,2}([0, T] \times \mathbb{R}^N; \mathbb{R})$ *satisfies*

$$\big|f(t, y)\big|_{\mathbb{R}^N} + \left|\frac{\partial f}{\partial t}(t, y) + \big[\mathcal{L}f(t, \cdot\,)\big](y)\right| \le K \exp\left(\frac{\epsilon|y|^2_{\mathbb{R}^N}}{2AT}\right)$$

for all $(t, y) \in [0, T] \times \mathbb{R}^N$ *and some* $K < \infty$ *and* $\epsilon \in (0, 1)$. *Then*

$$f\big(t \wedge T, p(t \wedge T, x, \mathbf{w})\big) - \int_0^{t\wedge T} \left(\frac{\partial f}{\partial t} + \mathcal{L}f\right)\big(\tau, p(\tau, x, \mathbf{w})\big)\, d\tau$$

is a $\mu_{\mathbb{R}^d}$-*martingale relative to* $\{\bar{\mathcal{B}}_t : t \ge 0\}$.

2.3.4. A Continuity Estimate. Just as (1.25) generalized to (2.43), so too (1.6) has an analog for the paths $p(\cdot, x, \mathbf{w})$. More precisely, by combining (2.39) with (1.6), we will show that, for each $\alpha \in \left(0, \frac{1}{2}\right)$ and $T \in (0, \infty)$, there exists a $C_\alpha(T) < \infty$ with the property that

$$\sup_{x \in \mathbb{R}^N} \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\left(\sup_{0 \le s < t \le T} \frac{\left|p(t, x, \mathbf{w}) - p(s, x, \mathbf{w})\right|_{\mathbb{R}^N}}{(t-s)^\alpha}\right)^2\right] \le C_\alpha(T). \tag{2.46}$$

In fact, we can more or less repeat the argument used in §1.1.3 . Namely, for given $0 \le s < t \le T$ with $t - s \le \frac{1}{2}$, determine $n \in \mathbb{N}$ by $2^{-n-2} \le t - s < 2^{-n-1}$. Then, for some $C \in [1, \infty)$,

$$\begin{aligned}
&\left|p(t, x, \mathbf{w}) - p(s, x, \mathbf{w})\right|_{\mathbb{R}^N} \\
&\quad \le \left|p_n(t, x, \mathbf{w}) - p_n(s, x, \mathbf{w})\right|_{\mathbb{R}^N} + 2 \sup_{\tau \in [0,T]} \left|p(\tau, x, \mathbf{w}) - p_n(\tau, x, \mathbf{w})\right|_{\mathbb{R}^N} \\
&\quad \le C \sup_{\tau \in [s,t]} \left|\mathbf{w}(\tau) - \mathbf{w}(s)\right|_{\mathbb{R}^d} + 2 \sup_{\tau \in [0,T]} \left|p(\tau, x, \mathbf{w}) - p_n(\tau, x, \mathbf{w})\right|_{\mathbb{R}^N} \\
&\quad \le C(t-s)^\alpha \left[\sup_{0 \le \sigma < \tau \le T} \frac{\left|\mathbf{w}(\tau) - \mathbf{w}(\sigma)\right|_{\mathbb{R}^d}}{(\tau - \sigma)^\alpha}\right. \\
&\qquad\qquad \left. + 2^{(n+2)\alpha + 1} \sup_{\tau \in [0,T]} \left|p(\tau, x, \mathbf{w}) - p_n(\tau, x, \mathbf{w})\right|_{\mathbb{R}^N}\right],
\end{aligned}$$

and so

$$\begin{aligned}
&\sup_{0 \le s < t \le T} \left(\frac{\left|p(t, x, \mathbf{w}) - p(s, x, \mathbf{w})\right|_{\mathbb{R}^N}}{(t-s)^\alpha}\right)^2 \\
&\quad \le 2C\left[\sup_{0 \le \sigma < \tau \le T} \left(\frac{\left|\mathbf{w}(\tau) - \mathbf{w}(\sigma)\right|_{\mathbb{R}^d}}{(\tau - \sigma)^\alpha}\right)^2\right. \\
&\qquad\qquad \left. + 64 \sum_{n=0}^{\infty} 2^{2n\alpha} \sup_{\tau \in [0,T]} \left|p(\tau, x, \mathbf{w}) - p_n(\tau, x, \mathbf{w})\right|_{\mathbb{R}^N}^2\right].
\end{aligned}$$

Thus, to complete the derivation of (2.46), all that we have to do is apply (1.6) and (2.39).

2.4 Derivatives

Although Theorem 2.40 shows that our approximation scheme works, it gives no information about $p(t, x, \mathbf{w})$ as a function of $x \in \mathbb{R}^N$. In fact, as yet, we do not even know that $p(t, x, \mathbf{w})$ is well-defined as a function of x: all that we know is that, for each x, $\mathbf{w} \rightsquigarrow p(\cdot, x, \mathbf{w})$ is well-defined up to a $\mu_{\mathbb{R}^d}$-null set (i.e., a set of $\mu_{\mathbb{R}^d}$-measure 0). On the other hand, each of our approximants $p_n(t, x, \mathbf{w})$ is a smooth function of x, and there is no reason to doubt that this smoothness survives in the limit. In fact, the purpose of this section is to check that the convergence in Theorem 2.40 can be promoted to local convergence in $C^{0,\infty}([0,\infty)\times\mathbb{R}^N;\mathbb{R}^N)$.

We begin by writing down the expressions on which our estimates will be based. Namely, define the Jacobian matrices (cf. (2.9) and (2.19))

$$J(\Xi,x)=\frac{\partial E}{\partial x}(\Xi,x)\quad\text{and}\quad \bar J(\Xi,x)=\frac{\partial \bar E}{\partial x}(\Xi,x)=J(\Xi,x)-I,\tag{2.47}$$

and set (cf. (2.17))

$$\bar J_{(m,n)}(t,x,\mathbf{w})=\bar J\big(\Xi_{(m,n)}(t,\mathbf{w}),p_n(T_{m,n},x,\mathbf{w})\big).\tag{2.48}$$

Then, from (2.17), we know that

$$\frac{\partial p_n}{\partial x}(t,x,\mathbf{w})-I=\sum_{m=0}^{\infty}\bar J_{(m,n)}(t,x,\mathbf{w})\frac{\partial p_n}{\partial x}(T_{m,n},x,\mathbf{w}).\tag{2.49}$$

More generally, given a multi-index $\beta\in\mathbb{N}^N$ with $|\beta|\ge 2$, use the chain rule to see that, for any $\mathbf{F}\in C^\infty(\mathbb{R}^N;\mathbb{R}^N)$:

$$\partial_x^\beta\mathbf{F}\circ E(\Xi,x)=J(\Xi,x)\big[\partial_x^\beta\mathbf{F}\big]\circ E(\Xi,x)+\sum_{\substack{\alpha<\beta\\|\alpha|\ge2}}\big[P_{\alpha,\beta}\mathbf{F}\big](x)\partial_x^\alpha E(\Xi,x),$$

where $[P_{\alpha,\beta}\mathbf{F}](x)$ is a universal $|\beta|$th order polynomial in $\{\partial_x^\gamma\mathbf{F}(x):\gamma<\alpha\}$. Hence, for $|\beta|\ge 2$,

$$\begin{aligned}\partial_x^\beta p_n(t,x,\mathbf{w})&=\sum_{m<2^nt}\bar J_{(m,n)}(t,x,\mathbf{w})\partial_x^\beta p_n(t,x,\mathbf{w})\\&\qquad+\sum_{\substack{\alpha<\beta\\|\alpha|\ge2}}\big[P_{\alpha,\beta}p_n(T_{m,n},\,*\,,\mathbf{w})\big](x)E^\alpha_{(m,n)}(t,x,\mathbf{w}),\end{aligned}\tag{2.50}$$

where

$$E^\alpha_{(m,n)}(t,x,\mathbf{w})\equiv\big[\partial_x^\alpha E\big(\Xi_{(m,n)}(t,\mathbf{w}),\,*\,\big)\big]\big(p_n(T_{m,n},x,\mathbf{w})\big).$$

2.4.1. Burkholder's Inequality. Because (2.50) involves polynomial expressions in the spacial derivatives of $p_n(\,\cdot\,,\,*\,,\mathbf{w})$, it is inevitable that estimates on any moments will require estimates on all moments of these derivatives. The technical device with which we will develop such estimates is the following variant of Burkholder's inequality, which is, in turn, a variant of Doob's Inequality.

2.51 Lemma. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\{\mathcal{F}_t : t \in [0,\infty)\}$ and $\{\mathcal{F}_m : m \in \mathbb{N}\}$ be non-decreasing families of sub σ-algebras of $\mathcal{F}$. Further, assume that, for each $m \in \mathbb{N}$, $(M_m(t), \mathcal{F}_t, \mathbb{P})$ is a right-continuous martingale with values in the separable Hilbert space H and $M_m(0) = \mathbf{0}$. Finally, for each $T \in [0,\infty)$, assume that $(\sum_{\ell=0}^m M_\ell(T), \mathcal{F}_m, \mathbb{P})$ is a martingale. Then, for each $q \in [1,\infty)$ and $T \in [0,\infty)$,*

$$\left\| \sup_{t\in[0,T]} \left\| \sum_{m=0}^\infty M_m(t) \right\|_H \right\|_{L^{2q}(\mathbb{P};\mathbb{R})} \le 2q \left(\sum_{m=0}^\infty \|M_m(T)\|^2_{L^{2q}(\mathbb{P};H)} \right)^{\frac{1}{2}}.$$

Proof: Obviously, it suffices for us to treat the case in which $M_m \equiv \mathbf{0}$ for all but a finite number of m's, in which case $M(t) \equiv \left\| \sum_{m=0}^\infty M_m(t) \right\|_H$ is a $\mathbb{P}$-submartingale relative to $\{\mathcal{F}_t : t \in [0,\infty)\}$. Thus, by Doob's Inequality,

$$\left\| \sup_{t\in[0,T]} \left\| \sum_{m=0}^\infty M_m(t) \right\|_H \right\|_{L^{2q}(\mathbb{P};\mathbb{R})} \le \frac{2q}{2q-1} \left\| \sum_{m=0}^\infty M_m(T) \right\|_{L^{2q}(\mathbb{P};H)}.$$

At the same time, by Burkholder's inequality (cf. Corollary 6.3.18 in [35]),

$$\left\| \sum_{m=0}^\infty M_m(T) \right\|_{L^{2q}(\mathbb{P};H)} \le (2q-1) \left\| \sum_{m=0}^\infty \|M_m(T)\|_H^2 \right\|^{\frac{1}{2}}_{L^q(\mathbb{P};\mathbb{R})}.$$

Hence, the desired estimate follows after an application of Minkowski's inequality. □

2.4.2. Estimating Derivatives. We begin by showing that, for each $q \in [1,\infty)$, there is a $C_q < \infty$ such that[12]

$$\left\| \sup_{t\in[0,T]} \left\| \frac{\partial p_n}{\partial x}(t,x,\mathbf{w}) \right\|_{\mathrm{H.S.}} \right\|_{L^{2q}(\mathbb{P};\mathbb{R})} \le C_q e^{C_q T} \tag{2.52}$$

for all $T \in [0,\infty)$. Namely, we start from (2.49) and, just as we did in §2.3.2, we begin by splitting off the martingale part. That is, set

$$M_{(m,n)}(t,x,\mathbf{w}) = \sum_{k=1}^d \xi^k_{(m,n)}(t,\mathbf{w}) \frac{\partial X_k}{\partial x}\big(p_n(T_{m,n},x,\mathbf{w})\big),$$

and write

$$\begin{aligned} \frac{\partial p_n}{\partial x}(t,x,\mathbf{w}) - I = \sum_{m=0}^\infty M_{(m,n)}(t,x,\mathbf{w}) \frac{\partial p_n}{\partial x}(T_{m,n},x,\mathbf{w}) \\ + \sum_{m=0}^\infty B_{(m,n)}(t,x,\mathbf{w}) \frac{\partial p_n}{\partial x}(T_{m,n},x,\mathbf{w}), \end{aligned}$$

[12] When A is a matrix, we use $\|A\|_{\mathrm{H.S.}}$ to denote its Hilbert-Schmidt norm.

where $B_{(m,n)} \equiv \bar{J}_{(m,n)} - M_{(m,n)}$. Next, by Lemma 2.51 and Wiener scaling, we see that, for each $q \in [1,\infty)$:

$$\begin{aligned}
&\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]}\left\|\sum_{m=0}^{\infty} M_{(m,n)}(t,x,\mathbf{w})\frac{\partial p_n}{\partial x}(t,x,\mathbf{w})\right\|_{\mathrm{H.S.}}^{2q}\right]\\
&\quad\le (2q)^{2q}\left(\sum_{m<2^nT}\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\left\|M_{(m,n)}(T,x,\mathbf{w})\frac{\partial p_n}{\partial x}(T_{m,n},x,\mathbf{w})\right\|_{\mathrm{H.S.}}^{2q}\right]^{\frac{1}{q}}\right)^{q}\\
&\qquad\le C_q(1+T)^{q-1}2^{-n}\sum_{m<2^nT}\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\left\|\frac{\partial p_n}{\partial x}(T_{m,n},x,\mathbf{w})\right\|_{\mathrm{H.S.}}^{2q}\right]
\end{aligned}$$

for some $C_q < \infty$ depending only on q and the bounds on the first derivatives of the X_k's. At the same time, by (2.19), there is another $C_q' < \infty$, depending on q and the bounds on the X_k's and error terms in (2.19), such that

$$\begin{aligned}
&\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]}\left\|\sum_{m=0}^{\infty} B_{(m,n)}(t,x,\mathbf{w})\frac{\partial p_n}{\partial x}(T_{m,n},x,\mathbf{w})\right\|_{\mathrm{H.S.}}^{2q}\right]\\
&\le \left(1+2^nT\right)^{2q-1}\sum_{m<2^nT}\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]}\left\|B_{(m,n)}(t,x,\mathbf{w})\frac{\partial p_n}{\partial x}(T_{m,n},x,\mathbf{w})\right\|_{\mathrm{H.S.}}^{2q}\right]\\
&\le C_p'(1+T)^{2q-1}2^{-n}\sum_{m<2^nT}\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\left\|\frac{\partial p_n}{\partial x}(T_{m,n},x,\mathbf{w})\right\|_{\mathrm{H.S.}}^{2q}\right].
\end{aligned}$$

Hence, after combining these and using Lemma 2.23, we arrive at (2.52).

To get estimates on derivatives of higher order, we use an induction procedure based on (2.50). More precisely, we can show that, for each for each $q \in [1,\infty)$ and $\beta \in \mathbb{N}^N$ with $|\beta| \ge 1$, there exists a $C_{q,\beta} < \infty$ such that

$$\sup_{n\in\mathbb{N}}\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]}\left|\partial_x^\beta p_n(t,x,\mathbf{w})\right|_{\mathbb{R}^N}^{2q}\right] \le C_{q,\beta}e^{C_{q,\beta}T}. \tag{2.53}$$

Indeed, starting from (2.50), exactly the same argument as the one just used to prove (2.52) shows that, for $\beta \in \mathbb{N}^N$ with $|\beta| \ge 2$:

$$\begin{aligned}
&\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]}\left|\partial_x^\beta p_n(t,x,\mathbf{w})\right|_{\mathbb{R}^N}^{2q}\right]\\
&\quad\le A_{q,\beta}(1+T^{2q-1})2^{-n}\sum_{m<2^nT}\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\left|\partial_x^\beta p_n(T_{m,n},x,\mathbf{w})\right|_{\mathbb{R}^N}^{2q}\right]\\
&\qquad\quad + B_{q,\beta}(1+T^{2q-1})2^{-n}\sum_{\alpha<\beta}\sum_{m<2^nT}\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\left|\partial_x^\alpha p_n(T_{m,n},x,\mathbf{w})\right|_{\mathbb{R}^N}^{2q|\beta|}\right]
\end{aligned}$$

with appropriate choices of finite constants $A_{q,\beta}$ and $B_{q,\beta}$. Hence, if we assume (2.53) for β's with $|\beta| \le n$, then another application of Lemma 2.23 allows us to prove it for β's with $|\beta| = n+1$.

2.4.3. A Little Bit of Sobolev. Knowing Theorem 2.40 and (2.53), the proof that $p(t,x,\mathbf{w})$ can be chosen to be a smooth function of $x \in \mathbb{R}^N$ becomes a straight-forward application of basic Sobolev theory. To be precise, given a separable Hilbert space H and a compactly supported, smooth map $\mathbf{F} : \mathbb{R}^N \longrightarrow H$, define

$$\|\mathbf{F}\|_{2,H}^{(n)} = \left(\sum_{|\beta|\le n} \left\|\partial_x^\beta \mathbf{F}\right\|_{L^2(\mathbb{R}^N;H)}^2\right)^{\frac{1}{2}} \quad \text{for } n \in \mathbb{N}.$$

As an easy application of the Fourier transform, one finds that, for each $n \in \mathbb{N}$, there exist $A_n \in (0,\infty)$ and $B_n \in (0,\infty)$ such that

$$\|\mathbf{F}\|_{2,H}^{(n)} \le A_n \sqrt{\|\mathbf{F}\|_{2,H}^{(0)} \|\mathbf{F}\|_{2,H}^{(2n)}}$$

and, for each $\alpha \in \mathbb{N}^N$,

$$\sup_{x\in\mathbb{R}^N} \left\|\partial_x^\alpha \mathbf{F}(x)\right\|_H \le B_n \|\mathbf{F}\|_{2,H}^{(n)} \quad \text{when } n > |\alpha| + \frac{N}{2}.$$

Thus, there exists a $C_n \in (0,\infty)$ with the property that

$$\sup_{x\in\mathbb{R}^N} \left\|\partial_x^\alpha \mathbf{F}\right\|_H \le C_n \sqrt{\|\mathbf{F}\|_{2,H}^{(0)} \|\mathbf{F}\|_{2,H}^{(2n)}} \quad \text{for } n > |\alpha| + \frac{N}{2}. \tag{2.54}$$

2.55 LEMMA. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and*

$$\omega \in \Omega \longmapsto \mathbf{Y}(\,\cdot\,, *, \omega) \in C^\infty\big([0,\infty) \times \mathbb{R}^N; H\big)$$

a measurable map. Then, for each $n > \frac{N}{2}$, there exists a $K_n \in (0,\infty)$ with the property that

$$\begin{aligned}
&\left\| \sup_{t\in[0,T]} \sup_{|x|\le R} \left\|\partial_x^\alpha \mathbf{Y}(t, *)\right\|_H \right\|_{L^{2q}(\mathbb{P};\mathbb{R})}^2 \\
&\qquad \le K_n R^N \sup_{|x|\le 2R} \left\| \sup_{t\in[0,T]} \|\mathbf{Y}(t,x)\|_H \right\|_{L^{2q}(\mathbb{P};\mathbb{R})} \\
&\qquad\qquad \times \sum_{|\beta|\le 2n} \sup_{|x|\le 2R} \left\| \sup_{t\in[0,T]} \|\partial_x^\beta \mathbf{Y}(t,x)\|_H \right\|_{L^{2q}(\mathbb{P};\mathbb{R})}
\end{aligned} \tag{2.56}$$

for all $R \in [1,\infty)$, $q \in [1,\infty)$, and $\alpha \in \mathbb{N}^N$ with $n > |\alpha| + \frac{N}{2}$.

PROOF: Choose $\psi \in C^\infty(\mathbb{R}^N;[0,1])$ so that

$$\psi(x) = \begin{cases} 1 & \text{if } |x| \le 1 \\ 0 & \text{if } |x| \ge 2, \end{cases}$$

and set $\mathbf{Y}_R(t,x,\omega) = \psi\left(\frac{x}{R}\right)\mathbf{Y}(t,x,\omega)$. Then, by (2.54) and Schwarz's inequality,

$$\left\|\sup_{t\in[0,T]}\sup_{|x|\le R}\left\|\partial_x^\alpha \mathbf{Y}(t,x)\right\|_H\right\|^2_{L^{2q}(\mathbb{P};\mathbb{R})}$$
$$\le C_n^2\left\|\sup_{t\in[0,T]}\|\mathbf{Y}_R(t,\cdot)\|^{(0)}_{2,H}\right\|_{L^{2q}(\mathbb{P};\mathbb{R})}\left\|\sup_{t\in[0,T]}\|\mathbf{Y}_R(t,\cdot)\|^{(2n)}_{2,H}\right\|_{L^{2q}(\mathbb{P};\mathbb{R})}.$$

At the same time, by the continuous form of Minkowski's inequality[13],

$$\left\|\sup_{t\in[0,T]}\|\mathbf{Y}_R(t,\cdot)\|^{(0)}_{2,H}\right\|_{L^{2q}(\mathbb{P};\mathbb{R})} \le \left(\int_{\mathbb{R}^N}\left\|\sup_{t\in[0,T]}\|\mathbf{Y}_R(t,x)\|_H\right\|^2_{L^{2q}(\mathbb{P};\mathbb{R})}dx\right)^{\frac{1}{2}}$$

and

$$\left\|\sup_{t\in[0,T]}\|\mathbf{Y}_R(t,\cdot)\|^{(2n)}_{2,H}\right\|_{L^{2q}(\mathbb{P};\mathbb{R})}$$
$$\le\left(\sum_{|\beta|\le 2n}\int_{\mathbb{R}^N}\left\|\sup_{t\in[0,T]}\|\partial_x^\beta\mathbf{Y}_R(t,x)\|_H\right\|^2_{L^{2q}(\mathbb{P};\mathbb{R})}dx\right)^{\frac{1}{2}};$$

from which (2.56) is an easy step. □

As an essentially immediate corollary of Lemma 2.55, we have the following useful criterion for uniform convergence of stochastic processes.

2.57 THEOREM. *Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and H a separable Hilbert space. Further, for each $n\in\mathbb{N}$, let*

$$\omega\in\Omega\longmapsto\mathbf{X}_n(\cdot,*,\omega)\in C^{0,\infty}\big([0,\infty)\times\mathbb{R}^N;H\big)$$

be a measurable map. Finally, assume that, for some $q\in[1,\infty)$ and all $T\in[0,\infty)$:

$$\sum_{n=0}^\infty\sup_{x\in\mathbb{R}^N}\left\|\sup_{t\in[0,T]}\left\|\mathbf{X}_{n+1}(t,x)-\mathbf{X}_n(t,x)\right\|_H\right\|_{L^{2q}(\mathbb{P};\mathbb{R})}<\infty$$

and

$$\sup_{n\in\mathbb{N}}\sup_{x\in\mathbb{R}^N}\left\|\sup_{t\in[0,T]}\left\|\partial_x^\alpha\mathbf{X}_n(t,x)\right\|_H\right\|_{L^{2q}(\mathbb{P};\mathbb{R})}<\infty,\quad\alpha\in\mathbb{N}^N.$$

---

[13] This is the form of Minkowski's inequality which says that, for any σ-finite measure spaces $(E_i,\mathcal{F}_i,\mu_i)$, $i\in\{1,2\}$, and any pair $1\le p_1\le p_2\le\infty$, one has: $\big\|\|f(\cdot_1,\cdot_2)\|_{L^{p_1}(\mu_1)}\big\|_{L^{p_2}(\mu_2)}\le\big\|\|f(\cdot_1,\cdot_2)\|_{L^{p_2}(\mu_2)}\big\|_{L^{p_1}(\mu_1)}$. See, for example, Theorem 6.2.14 in [36].

Then there exists a measurable

$$\omega \in \Omega \longmapsto \mathbf{X}(\,\cdot\,,\,*\,,\omega) \in C^{0,\infty}\big([0,\infty)\times\mathbb{R}^N;H\big)$$

with the property that, for each $T \in (0,\infty)$, $R \in [0,\infty)$, and $\alpha \in \mathbb{N}^N$:

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \sup_{|x|\le R} \left\|\partial_x^\alpha \mathbf{X}_n(t,\,*\,) - \partial_x^\alpha \mathbf{X}(t,\,*\,)\right\|_H = 0$$

both $\mathbb{P}$-almost surely and in $L^{2q}(\mathbb{P};\mathbb{R})$.

2.4.4. Existence of a Smooth Choice. By combining the estimates in (2.38), Theorem 2.42, and (2.53) with Theorem 2.57, we get the following improvement of our original statement.

2.58 THEOREM. *The family of maps $\mathbf{w} \rightsquigarrow p(\,\cdot\,,x,\mathbf{w})$ as x runs over $\mathbb{R}^N$ admit a $\mu_{\mathbb{R}^d}$-equivalent version*[14] *as a measurable map*

$$\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto p(\,\cdot\,,\,*\,,\mathbf{w}) \in C^{0,\infty}\big([0,\infty)\times\mathbb{R}^N;\mathbb{R}^N\big).$$

In fact, for each $\beta \in \mathbb{N}^N$, $T \in (0,\infty)$, and $R \in (0,\infty)$,

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \sup_{|x|_{\mathbb{R}^N}\le R} \left|\partial_x^\beta p_n(t,x,\mathbf{w}) - \partial_x^\beta p(t,x,\mathbf{w})\right|_{\mathbb{R}^N} = 0$$

both $\mu_{\mathbb{R}^d}$-almost surely and in $L^2(\mu_{\mathbb{R}^d};\mathbb{R})$. Moreover, for each $q \in [1,\infty)$, $\beta \in \mathbb{N}^N$, and $R \in (0,\infty)$, there is a $C_{q,\beta}(R) < \infty$ for which

$$\sup_{x\in\mathbb{R}^N} \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]} \sup_{|y-x|_{\mathbb{R}^N}\le R} \left|\partial_y^\beta p(t,y,\mathbf{w})\right|_{\mathbb{R}^N}^q\right]^{\frac{1}{q}} \le C_{q,\beta}(R)e^{C_{q,\beta}(R)T}.$$

In particular, if $f \in C_b^\infty(\mathbb{R}^N;\mathbb{R})$, then

$$(t,x) \in [0,\infty)\times\mathbb{R}^N \longmapsto \mathbb{E}^{\mu_{\mathbb{R}^d}}\big[f\circ p(t,x,\mathbf{w})\big] \in \mathbb{R}$$

is an element of $C^{0,\infty}([0,\infty)\times\mathbb{R}^N;\mathbb{R})$. In fact, for each $n \in \mathbb{N}$, there is a $C_n < \infty$ such that

$$\max_{|\beta|\le n} \sup_{\substack{t\in[0,T]\\ x\in\mathbb{R}^N}} \left|\partial_x^\beta \mathbb{E}^{\mu_{\mathbb{R}^d}}\big[f\circ p(t,x,\mathbf{w})\big]\right| \le C_n e^{C_n T}\|f\|_{C_b^n([0,T]\times\mathbb{R}^N;\mathbb{R})}.$$

[14] For those who are not familiar with the technical difficulties of dealing with uncountable families of random variables, like $\{p(\,\cdot\,,x,\mathbf{w}) : x \in \mathbb{R}^N\}$, this notion may be strange. The meaning is that the new version, say $\{q(\,\cdot\,,x,\mathbf{w}) : x \in \mathbb{R}^N\}$, has the described property and, for each $x \in \mathbb{R}^N$, agrees $\mu_{\mathbb{R}^d}$-almost surely with the old. In the case at hand, $q(\,\cdot\,,\,*\,,\mathbf{w}) \in C^{0,\infty}([0,\infty)\times\mathbb{R}^N;\mathbb{R}^N)$ for each $\mathbf{w}$ and $q(\,\cdot\,,x,\mathbf{w}) = p(\,\cdot\,,x,\mathbf{w})$ (a.s., $\mu_{\mathbb{R}^d}$) for each $x \in \mathbb{R}^N$.

2.4.5. Loosening Things Up. The existence statements in Theorems 2.40 and 2.58 were the main goal of this section. However, it is important to notice that the approximation scheme is more flexible than we have indicated heretofore. In particular, there is nothing sacrosanct about the interpolation times $\{T_{m,n} : m \in \mathbb{N}\}_{n=0}^{\infty}$. In fact, they can be replaced by any sequence $\{\mathcal{T}_n\}_{n=0}^{\infty}$ where, for each $n \in \mathbb{N}$, $\mathcal{T}_n$ is a non-decreasing family $\{\tau_{m,n} : m \in \mathbb{N}\}$ of $\{\bar{\mathcal{B}}_t : t \geq 0\}$-stopping times[15] with the properties that

$$\tau_{0,n}(\mathbf{w}) = 0 \quad \text{and} \quad (m-1)2^{-n} \leq \tau_{m,n}(\mathbf{w}) \leq m2^{-n} \quad \text{for } m \geq 1.$$

That is, if

$$\begin{aligned}\Xi_{m,\mathcal{T}_n}(t,\mathbf{w}) &= \left(\xi^0_{m,\mathcal{T}_n}(t,\mathbf{w}), \boldsymbol{\xi}_{m,\mathcal{T}_n}(t,\mathbf{w})\right)\\ &\equiv \Big(t\wedge\tau_{m+1,n}(\mathbf{w}) - t\wedge\tau_{m,n}(\mathbf{w}), \mathbf{w}\big(t\wedge\tau_{m+1,n}(\mathbf{w})\big) - \mathbf{w}\big(t\wedge\tau_{m,n}(\mathbf{w})\big)\Big),\end{aligned}$$

and $\mathbf{w}\rightsquigarrow p_{\mathcal{T}_n}(\,\cdot\,,x,\mathbf{w})$ is defined by

$$\begin{aligned} p_{\mathcal{T}_n}(0,x,\mathbf{w}) &= x \quad \text{and}\\ p_{\mathcal{T}_n}(t,x,\mathbf{w}) &= E\Big(\Xi_{m,\mathcal{T}_n}(t,\mathbf{w}), p_{\mathcal{T}_n}\big(\tau_{m,n}(\mathbf{w}),x,\mathbf{w}\big)\Big)\end{aligned}$$

for $t \in \big[\tau_{m,n}(\mathbf{w}), \tau_{m+1,n}(\mathbf{w})\big]$, then, for each $\beta \in \mathbb{N}^N$ and $R \in (0,\infty)$,

$$\begin{aligned}\lim_{n\to\infty}\sup_{t\in[0,T]}\sup_{|x|_{\mathbb{R}^N}\leq R}\left|\partial_x^\beta p_{\mathcal{T}_n}(t,x,\mathbf{w}) - \partial_x^\beta p(t,x,\mathbf{w})\right|_{\mathbb{R}^N} = 0\\ \text{both } \mu_{\mathbb{R}^d}\text{-almost surely and in } L^2(\mu_{\mathbb{R}^d};\mathbb{R}).\end{aligned} \tag{2.59}$$

The proof of (2.59) relies on the *strong independent increment property* (cf. Theorem 4.3.3 and Exercise 4.3.45 in [35]) which says that, for any $\{\bar{\mathcal{B}}_t : t \geq 0\}$-stopping time $\sigma : \mathfrak{W}(\mathbb{R}^d) \longrightarrow [0,\infty]$, the $\mu_{\mathbb{R}^d}$-distribution of

$$\mathbf{w} \in \{\sigma < \infty\} \longmapsto \Big(\mathbf{w}\big(\,\cdot\,\wedge\sigma(\mathbf{w})\big), \delta_{\sigma(\mathbf{w})}\mathbf{w}\Big) \in \mathfrak{W}(\mathbb{R}^d)\times\mathfrak{W}(\mathbb{R}^d),$$

where

$$\big[\delta_s\mathbf{w}\big](t) \equiv \mathbf{w}(s+t) - \mathbf{w}(s) \text{ for } (s,t) \in [0,\infty)^2, \tag{2.60}$$

is the same as that of

$$(\mathbf{w}_1,\mathbf{w}_2) \in \{\sigma<\infty\}\times\mathfrak{W}(\mathbb{R}^d) \longmapsto \Big(\mathbf{w}_1\big(\,\cdot\,\wedge\sigma(\mathbf{w}_1)\big), \mathbf{w}_2\Big) \in \mathfrak{W}(\mathbb{R}^d)\times\mathfrak{W}(\mathbb{R}^d)$$

under $\mu_{\mathbb{R}^d}\times\mu_{\mathbb{R}^d}$. Knowing this, it is an easy matter to repeat the argument leading to (2.53) and thereby conclude that the same estimate holds after

[15] A function $\sigma : \mathfrak{W}(\mathbb{R}^d) \longrightarrow [0,\infty]$ is said to be a $\{\bar{\mathcal{B}}_t : t > 0\}$-stopping time (cf. the beginning of §4.3 in [35]) if $\{\sigma \leq t\} \in \bar{\mathcal{B}}_t$ for each $t \in [0,\infty)$.

$p_n(t,x,\mathbf{w})$ has been replaced by $p_{\mathcal{T}_n}(t,x,\mathbf{w})$. Thus, all that remains is to show that $p_n(t,x,\mathbf{w})$ can be replaced by $p_{\mathcal{T}_n}(t,x,\mathbf{w})$ in the first part of Theorem 2.40. To this end, take $\sigma_{2m,n} \equiv m2^{-n}$ and $\sigma_{2m+1,n} \equiv \tau_{m+1,n}$, set $\mathcal{S}_n = \{\sigma_{m,n} : m \in \mathbb{N}\}$, and define $p_{\mathcal{S}_n}(t,x,\mathbf{w})$ accordingly. Then exactly the same arguments which led to (2.38) show that, for some $C < \infty$,

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]} \left|p_n(t,x,\mathbf{w}) - p_{\mathcal{S}_n}(t,x,\mathbf{w})\right|^2_{\mathbb{R}^N} \vee \left|p_{\mathcal{T}_n}(t,x,\mathbf{w}) - p_{\mathcal{S}_n}(t,x,\mathbf{w})\right|^2_{\mathbb{R}^N}\right] \le Ce^{CT}2^{-n}.$$

Hence, the desired result for the general case is, in fact, a corollary of original result for $\{p_n(\,\cdot\,,*,\mathbf{w})\}_{n=0}^\infty$.

2.5 The Flow Property

Perhaps the most important application of the considerations in §2.4 is that they allow us to discuss and prove flow properties for the map $p(t,x,\mathbf{w})$. As a consequence, we will be able to prove that the family $\{\mathbf{P}_t^{\mathcal{L}} : t > 0\}$ of operators given by

$$\left[\mathbf{P}_t^{\mathcal{L}} f\right](x) = \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[f \circ p(t,x,\mathbf{w})\right], \quad f \in C_{\mathrm{b}}(\mathbb{R}^N;\mathbb{R}) \tag{2.61}$$

form a contraction semigroup. Once we know this, it will be easy to see that, when $f \in C^2_{\mathrm{b}}(\mathbb{R}^N;\mathbb{R})$, $(t,x)\rightsquigarrow[\mathbf{P}_t^{\mathcal{L}} f](x)$ solves the Cauchy initial value problem

$$\frac{\partial u}{\partial t}(t,x) = \left[\mathcal{L}u(t,*)\right](tx) \quad \text{in } [0,\infty)\times\mathbb{R}^N \text{ with } u(0,\,\cdot\,) = f, \tag{2.62}$$

and this, in turn, will allow us to show that, for each $x \in \mathbb{R}^N$, the distribution $\mathbb{P}_x^{\mathcal{L}}$ of $\mathbf{w}\rightsquigarrow p(\,\cdot\,,x,\mathbf{w})$ under $\mu_{\mathbb{R}^d}$ is the one and only solution to the martingale problem for $\mathcal{L}$ starting at x.

2.5.1. Renewal at Stopping Times. Our goal here is to prove that (cf. (2.60))

$$p\big(t+\sigma(\mathbf{w}),x,\mathbf{w}\big) = p\Big(t, p\big(\sigma(\mathbf{w}),x,\mathbf{w}\big), \delta_{\sigma(\mathbf{w})}\mathbf{w}\Big), \quad t\in[0,\infty), \quad (\text{a.s.}, \mu_{\mathbb{R}^d}) \text{ on } \{\sigma<\infty\} \tag{2.63}$$

for every $\{\bar{\mathcal{B}}_t : t > 0\}$-stopping time σ.[16]

To prove (2.63), first note that it suffices to treat the case in which $\sigma \le T$ for some $T \in (0,\infty)$. Indeed, if this is not the case, one can replace σ by $\sigma \wedge T$

[16] For the benefit of those readers who are sensitive enough to be concerned about such niceties, it should be pointed out that the only reason why (2.63) "makes sense" is the strong independent increment property, alluded to in §2.4.5.

and then pass to limit as $T \nearrow \infty$. Thus, we will assume that $\sigma \le T$. Next, for $n \in \mathbb{N}$, take $\mathcal{T}_n = \{\tau_{m,n} : m \in \mathbb{N}\}$ where $\tau_{0,n} \equiv 0$ and

$$\tau_{m+1,n}(\mathbf{w}) = \begin{cases} \left(\tau_{m,n}(\mathbf{w}) + 2^{-n}\right) \wedge \sigma(\mathbf{w}) & \text{if } \tau_{m,n}(\mathbf{w}) \le \sigma(\mathbf{w}) \\ \tau_{m,n}(\mathbf{w}) + 2^{-n} & \text{otherwise.} \end{cases}$$

Then (cf. §2.4.5)

$$p_{\mathcal{T}_n}\big(t + \sigma(\mathbf{w}), x, \mathbf{w}\big) = p_{\mathcal{T}_n}\Big(t, p_{\mathcal{T}_n}\big(\sigma(\mathbf{w}), x, \mathbf{w}\big), \delta_{\sigma(\mathbf{w})}\mathbf{w}\Big)$$

for every $(t, \mathbf{w}) \in [0, \infty) \times \mathfrak{W}(\mathbb{R}^d)$. Hence, by the strong independent increment property combined with (2.59), we get (2.63) by letting $n \to \infty$.

As a consequence of (2.63), we obtain the following *strong Markov property* for the $\mu_{\mathbb{R}^d}$-distribution of $\mathbf{w} \rightsquigarrow p(\,\cdot\,, x, \mathbf{w})$. Namely, let $\bar{\mathcal{B}}_\sigma$ be the σ-algebra over $\mathfrak{W}(\mathbb{R}^d)$ corresponding to the $\{\bar{\mathcal{B}}_t : t \ge 0\}$-stopping time σ.[17] Then, by (2.63) combined with the strong increment property:

$$\begin{aligned} &\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\mathbf{1}_{[0,\infty)}\big(\sigma(\mathbf{w})\big) F \circ p(\,\cdot\, + \sigma(\mathbf{w}), x, \mathbf{w}) \,\Big|\, \bar{\mathcal{B}}_\sigma\Big](\mathbf{w}') \\ (2.64)\qquad &= \mathbf{1}_{[0,\infty)}\big(\sigma(\mathbf{w}')\big) \int_{\mathfrak{W}(\mathbb{R}^d)} F \circ p\Big(\,\cdot\,, p\big(\sigma(\mathbf{w}'), x, \mathbf{w}'\big), \mathbf{w}\Big)\, \mu_{\mathbb{R}^d}(d\mathbf{w}) \\ &\qquad\qquad \text{for } \mu_{\mathbb{R}^d}\text{-almost every } \mathbf{w}' \in \mathfrak{W}(\mathbb{R}^d), \end{aligned}$$

whenever $F : \mathcal{P}(\mathbb{R}^N) \longrightarrow \mathbb{R}$ is bounded and measurable.

2.5.2. The Heat Flow Semigroup for $\mathcal{L}$ and Uniqueness. Given a bounded, measurable $f : \mathbb{R}^N \longrightarrow \mathbb{R}$, define $\mathbf{P}_t^{\mathcal{L}} f$ as in (2.61). Clearly, for each $t \in [0, \infty)$, $\mathbf{P}_t^{\mathcal{L}}$ is a non-negative contraction operator on bounded measurable functions. In fact, by Theorem 2.58, for each $n \in \mathbb{N}$, $\mathbf{P}_t^{\mathcal{L}}$ preserves $C_{\mathrm{b}}^n(\mathbb{R}^N; \mathbb{R})$ and satisfies

$$(2.65)\qquad \big\|\mathbf{P}_t^{\mathcal{L}} f\big\|_{C_{\mathrm{b}}^n(\mathbb{R}^N;\mathbb{R})} \le C_n e^{C_n t} \|f\|_{C_{\mathrm{b}}^n(\mathbb{R}^N;\mathbb{R})}$$

for some $C_n < \infty$. Moreover, (2.64) for constant σ's says that $\{\mathbf{P}_t^{\mathcal{L}} : t \ge 0\}$ forms a semigroup in the sense that

$$(2.66)\qquad \mathbf{P}_{t_1+t_2}^{\mathcal{L}} = \mathbf{P}_{t_2}^{\mathcal{L}} \circ \mathbf{P}_{t_1}^{\mathcal{L}} \quad \text{for all } (t_1, t_2) \in [0, \infty)^2.$$

Finally, we want to show that if $f \in C_{\mathrm{b}}^2(\mathbb{R}^N; \mathbb{R})$, then the function $(t, x) \in [0, \infty) \rightsquigarrow [\mathbf{P}_t^{\mathcal{L}} f](x)$ solves (2.62); and clearly this reduces to checking that

$$(2.67)\qquad \frac{\partial}{\partial t}\big[\mathbf{P}_t^{\mathcal{L}} f\big](x) = \big[\mathcal{L}\mathbf{P}_t^{\mathcal{L}} f\big](x).$$

[17] That is, $\bar{\mathcal{B}}_\sigma$ is the set of $A \subseteq \mathfrak{W}(\mathbb{R}^d)$ such that $A \cap \{\sigma \le t\} \in \bar{\mathcal{B}}_t$ for all $t \in [0, \infty)$. Alternatively, by Exercise 4.3.45 in [35], $\bar{\mathcal{B}}_\sigma$ is the $\mu_{\mathbb{R}^d}$-completion of the σ-algebra generated by $\{\mathbf{w}(t \wedge \sigma(\mathbf{w})) : t \ge 0\}$.

To this end, first note that, by the last part of Theorem 2.40,

$$\left[\mathbf{P}_h^{\mathcal{L}} f\right](x) - f(x) = h[\mathcal{L}f](x) + o(h) \quad \text{as } h \searrow 0$$

for any $f \in C_{\mathrm{b}}^2(\mathbb{R}^N;\mathbb{R})$. Hence, by applying this to $\mathbf{P}_t^{\mathcal{L}} f$ and using (2.66), we see that

$$\lim_{h\searrow 0} \frac{[\mathbf{P}_{t+h}^{\mathcal{L}} f](x) - [\mathbf{P}_t^{\mathcal{L}} f](x)}{h} = \left[\mathcal{L}\mathbf{P}_t^{\mathcal{L}} f\right](x),$$

which is what we wanted to prove. Of course, after combining this with (2.65), we now know that if $f \in C_{\mathrm{b}}^{2n}(\mathbb{R}^N;\mathbb{R})$, then, for each $T \in [0,\infty)$, $(t,x) \in [0,T] \times \mathbb{R}^N \longmapsto [\mathbf{P}_t^{\mathcal{L}} f](x) \in \mathbb{R}$ is an element of $C_{\mathrm{b}}^{n,2n}([0,T]\times\mathbb{R}^N;\mathbb{R})$.

For us, the most significant application of the preceding is the following complement to the existence statement in Theorem 2.40.

2.68 THEOREM. *For each $x \in \mathbb{R}^N$, the $\mu_{\mathbb{R}^d}$-distribution of $\mathbf{w} \rightsquigarrow p(\,\cdot\,, x, \mathbf{w})$ is the one and only solution $\mathbb{P}_x^{\mathcal{L}}$ to the martingale problem for $\mathcal{L}$ starting at x.*

PROOF: The proof is essentially the same as that of the uniqueness assertion in Theorem 2.12. Namely, given a solution $\mathbb{P}$ to the martingale problem for $\mathcal{L}$ starting at x, we want to show that, for any $f \in C_{\mathrm{b}}^\infty(\mathbb{R}^N;\mathbb{R})$ and $0 \le t < T < \infty$,

$$\mathbb{E}^{\mathbb{P}}\left[f\big(p(T)\big)\,\middle|\,\mathcal{F}_t\right] = \left[\mathbf{P}_{T-t}^{\mathcal{L}} f\right]\big(p(t)\big) \quad (\text{a.s.}, \mathbb{P}). \tag{2.69}$$

Indeed, knowing this, the proof that $\mathbb{P}$ is unique follows from exactly the same inductive procedure as we outlined at the end of §2.2.2.

To prove (2.69), one proceeds, just as in the proofs of Theorem 1.24 and Corollary 2.42, to extend the class of test functions to include time-dependent functions which do not necessarily have compact support. That is, one shows that

$$g\big(t\wedge T, p(t\wedge T)\big) - \int_0^{t\wedge T} \left(\frac{\partial g}{\partial \tau} + \mathcal{L}g\right)\big(\tau, p(\tau)\big)\,d\tau$$

is a $\mathbb{P}$-martingale relative to $\{\mathcal{F}_t : t \ge 0\}$ whenever $g \in C_{\mathrm{b}}^\infty([0,T]\times\mathbb{R}^N;\mathbb{R})$. In particular, given $f \in C_{\mathrm{b}}^\infty(\mathbb{R}^N;\mathbb{R})$, apply this with $g(\tau,\,\cdot\,) = \mathbf{P}_{T-\tau}^{\mathcal{L}} f$, use (2.67), and conclude that

$$\begin{aligned}\left[\mathbf{P}_{T-t}^{\mathcal{L}} f\right]\big(p(t)\big) &= g\big(t, p(t)\big) = \mathbb{E}^{\mathbb{P}}\left[g\big(T, p(T)\big)\,\middle|\,\mathcal{F}_t\right] \\ &= \mathbb{E}^{\mathbb{P}}\left[f\big(p(T)\big)\,\middle|\,\mathcal{F}_t\right] \qquad (\text{a.s.}, \mathbb{P}). \quad \square\end{aligned}$$

CHAPTER 3

Some Addenda, Extensions, and Refinements

We continue with the program initiated in Chapter 2. That is, $\mathcal{L}$ is an operator on $C^\infty(\mathbb{R}^N;\mathbb{R})$ given by (2.3), and we are attempting to construct and analyze the associated diffusion on $\mathbb{R}^N$. Our basic results are those contained in Theorems 2.12, 2.40, 2.58, and 2.68. However, there are several questions that are important to ask about these results, and we will deal with some of them in this chapter.

3.1 Explosion and Non-explosion

In §2.2 we dealt with commuting vector fields $(X_0, \ldots, X_d)$ for which X_0 is forward complete and each X_k, $1 \le k \le d$, is complete. Under these conditions, we saw (cf. Theorem 2.12) that the associated diffusion could be constructed by simply evaluating the map $E(\,\cdot\,, x)$ along $t \rightsquigarrow (t, \mathbf{w}(t))$. In particular, the completeness assumptions about the vector fields were sufficient to guarantee that the associated diffusion *never explodes*. That is, the diffusion lives on $\mathbb{R}^N$ for all time. On the other hand, when we turned in §2.3 to the non-abelian case, we imposed the much more rigid condition that the X_k's and all their derivatives be bounded. One should ask to what extent this rigid condition can be softened, and we will begin by making sure that the completeness assumptions which worked in the Abelian case are not sufficient in the general case.

3.1.1. An Example. The following example shows that when the vector fields do not commute, then completeness fails to guarantee non-explosion.

Take $N = 3$, use (x, y, z) to denote a generic point in $\mathbb{R}^3$, set[1]

$$\Theta = y\partial_z - z\partial_y \quad \text{and} \quad \rho(x) = 1 + (x)^2,$$

and define

$$X_1 = y\partial_x - z\rho(x)\Theta \quad \text{and} \quad X_2 = z\partial_x + y\rho(x)\Theta$$

for $x = (x, y, z) \in \mathbb{R}^3$. Notice that, since $\Theta(y^2 + z^2) \equiv 0$, for any (x, y, z) and any $(\alpha_1, \alpha_2) \in \mathbb{R}^2$, the integral curve of $\alpha_1 X_1 + \alpha_2 X_2$ starting at (x, y, z) stays

[1] We use ∂_x to denote $\frac{\partial}{\partial x}$, etc.

on $\mathbb{R} \times \mathbf{S}^1(r)$, where $\mathbf{S}^1(r)$ is the circle of radius $r = \sqrt{y^2 + z^2}$. Hence, every vector field in the linear span of X_1 and X_2 is complete. Next, consider

$$\mathcal{L} \equiv \frac{1}{2}\left(X_1^2 + X_2^2\right) = \frac{y^2 + z^2}{2}\left(\partial_x^2 + \rho(x)\partial_x + \rho(x)^2\Theta^2\right),$$

and suppose that $\mathbb{P}$ were is a solution to the martingale problem for $\mathcal{L}$ starting at $(0, 1, 0)$. Since $\mathcal{L}f \equiv 0$ for any $f \in C^2(\mathbb{R}^3; \mathbb{R})$ which is a function of $y^2 + z^2$ alone, it is clear that $\mathbb{P}$ would have to satisfy

$$\mathbb{P}\Big(\big(p^2(t)\big)^2 + \big(p^3(t)\big)^2 = 1 \text{ for all } t \in [0, \infty)\Big) = 1.$$

Thus, for any $u \in C_c^2(\mathbb{R}; \mathbb{R})$,

$$u\big(p^1(t)\big) - \int_0^t \tfrac{1}{2}\Big(u''\big(p^1(s)\big) + \big(1 + p^1(s)^2\big)u'\big(p^1(s)\big)\Big)\, ds$$

would be a $\mathbb{P}$-martingale. In particular, if

$$R(\xi) = \int_0^{\xi} \left(1 + \eta^2\right) d\eta \quad \text{and} \quad u(\xi) = 2\int_0^{|\xi|} e^{-R(\zeta)} \left(\int_0^{\zeta} e^{R(\eta)}\, d\eta\right) d\zeta,$$

then, for any $L \in (0, \infty)$, we would have that

$$\mathbb{E}^{\mathbb{P}}\big[\zeta_L\big] = u(L), \quad \text{where } \zeta_L(p) \equiv \inf\big\{t \ge 0 : \left|p^1(t)\right| \ge L\big\}.$$

To see this, first note that $u''(\xi) + (1+\xi^2)u'(\xi) = 2$; and so, by Doob's Stopping Time Theorem, if $u_L \in C_c^2(\mathbb{R}; \mathbb{R})$ is chosen so that $u_L = u$ on $[-L, L]$, then

$$u\Big(p^1\big(t \wedge \zeta_L(p)\big)\Big) - t \wedge \zeta_L(p) = u_L\Big(p^1\big(t \wedge \zeta_L(p)\big)\Big) - t \wedge \zeta_L(p)$$

is a bounded $\mathbb{P}$-martingale which vanishes at $t = 0$. Since this would mean that

$$\mathbb{E}^{\mathbb{P}}\big[t \wedge \zeta_L(p)\big] = \mathbb{E}^{\mathbb{P}}\Big[u\big(t \wedge \zeta_L(p)\big)\Big] \quad \text{for all } t \ge 0,$$

the required conclusion follows after one lets $t \nearrow \infty$. But

$$\begin{aligned} u(\xi) &\le 2\int_0^{\infty} e^{R(\eta)} \left(\int_{\eta}^{\infty} e^{-R(\xi)}\, d\xi\right) d\eta \\ &= 2\int_0^{\infty} \frac{1}{1 + \eta^2}\, d\eta \\ &\qquad - 4\int_0^{\infty} e^{R(\eta)} \left(\int_{\eta}^{\infty} \frac{\xi}{(1 + \xi^2)^2} e^{-R(\xi)}\, d\xi\right) d\eta \le \pi; \end{aligned}$$

and so we would arrive at the conclusion that

$$\mathbb{E}^{\mathbb{P}}\big[\zeta_\infty(p)\big] \le \pi \quad \text{where } \zeta_\infty(p) \equiv \inf\{t \ge 0 : |p^1(t)| = \infty\}.$$

In other words, $\mathbb{P}$-almost every path would *explode* in a finite amount of time. Since this conclusion obviously contradicts the assumption that $\mathbb{P}$ lives on $\mathcal{P}(\mathbb{R}^3)$, we have now shown that *there is no solution to the martingale problem for $\mathcal{L}$ starting at* $(0,1,0)$.

3.2 Localization

As the preceding example demonstrates, explosion represents a possibility whose occurrence is difficult to predict from properties of the classical flows generated by the the vector fields $(X_0, X_1, \dots, X_d)$. On the other hand, as long as explosion does not occur, nothing goes wrong with the basic conclusions which we reached in Chapter 2. That is, just as in the theory of ordinary differential equations, questions about both existence and uniqueness are inherently local problems, as distinguished from explosion, which is an inherently global problem.

To understand this comment, first recall what one does in the setting of ordinary differential equations. Thus, let X be a smooth vector field on $\mathbb{R}^N$, and consider the problem of constructing the integral curve of X starting at some $x \in \mathbb{R}^N$. One way to proceed is to take an increasing exhaustion $\{U_m : m \ge 1\}$ of $\mathbb{R}^N$ by bounded open sets, choose $\psi_m \in C_c^\infty(\mathbb{R}^N;[0,1])$ so that $\psi_m \equiv 1$ on U_m, and let $p^m(\,\cdot\,,x)$ be the integral curve of $\psi_m X$ starting at x. If $\zeta^m(x) = \inf\{t \ge 0 : p^m(t,x) \notin U_m\}$, then

$$p^{m+1}(\,\cdot\,,x) \restriction \big[0,\zeta^m(x)\big) = p^m(\,\cdot\,,x) \restriction \big[0,\zeta^m(x)\big).$$

In particular, $\zeta^m(x) \le \zeta^{m+1}(x)$, and so the *explosion time*

$$\mathfrak{e}(x) \equiv \lim_{m\to\infty} \zeta^m(x)$$

exists. Moreover, an easy argument shows that $\mathfrak{e}(x)$ is independent of the choice of exhaustion. In addition, there exists a well-defined $t \in \big[0,\mathfrak{e}(x)\big) \longmapsto p(t,x) \in \mathbb{R}^N$ such that $p(\,\cdot\,,x) \restriction \big[0,\zeta^m(x)\big) = p^m(\,\cdot\,,x) \restriction \big[0,\zeta^m(x)\big)$ for each m. In fact, if we use $\widehat{\mathbb{R}^N} \equiv \mathbb{R}^N \cup \{\infty\}$ to denote the one-point compactification of $\mathbb{R}^N$, then, even if $\mathfrak{e}(x) < \infty$, we can extend $p(\,\cdot\,,x)$ as a continuous path from $[0,\infty)$ into $\widehat{\mathbb{R}^N}$ by setting $p(t,x) = \infty$ when $\mathfrak{e}(x) \le t < \infty$. Finally, observe that $x \rightsquigarrow \zeta^m(x)$ is a lower semi-continuous function for each m, and therefore so is $x \rightsquigarrow \mathfrak{e}(x)$. At the same time, if $\mathfrak{e}(x) < \infty$ and $x_n \to x$, then it is clear that $\overline{\lim}_{n\to\infty} \zeta^m(x_n) \le \mathfrak{e}(x)$ for every m. Hence, $\overline{\lim}_{n\to\infty} \mathfrak{e}(x_n) \le \mathfrak{e}(x)$. That is, $x \in \mathbb{R}^N \longmapsto e(x) \in [0,\infty]$ is continuous, from which it is an easy step to the conclusion that $x \in \mathbb{R}^N \longmapsto p(\,\cdot\,,x) \in C\big([0,\infty);\widehat{\mathbb{R}^N}\big)$ is also continuous.

It is our purpose now to carry out the corresponding line of reasoning in the stochastic setting.

3.2.1. Random Paths which may Explode. Given an open set U in $\mathbb{R}^N$, define the *first exit time* $\zeta^U : \mathcal{P}(\mathbb{R}^N) \longrightarrow [0,\infty]$ so that

$$\zeta^U(p) = \inf\{t \geq 0 : p(t) \notin U\}, \tag{3.1}$$

and observe that $p \rightsquigarrow \zeta^U(p)$ is lower semi-continuous. Next, let $\{U_m : m \geq 1\}$ be an increasing exhaustion of $\mathbb{R}^N$ by bounded open sets, and choose $\{\psi_m : m \geq 1\} \subseteq C_c^\infty(\mathbb{R}^N;[0,1])$ so that $\psi_m \equiv 1$ on an open neighborhood of $\bar{U}_m$. Finally, given a family $(X_0,\dots,X_d)$ of smooth vector fields on $\mathbb{R}^N$, set

$$X_{k,m} = \psi_m X_k \quad \text{and} \quad \mathcal{L}_m = X_{0,m} + \frac{1}{2}\sum_{k=1}^d X_{k,m}^2, \tag{3.2}$$

and, for each m, let

$$\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto p^m(\,\cdot\,, *, \mathbf{w}) \in C^{0,\infty}\big([0,\infty) \times \mathbb{R}^N;\mathbb{R}\big)$$

be the map (guaranteed by Theorems 2.40 and 2.58) corresponding to the vector fields $(X_{0,m},\dots,X_{d,m})$. Finally, define

$$\zeta^m(x,\mathbf{w}) = \zeta^{U_m}\big(p^m(\,\cdot\,,x,\mathbf{w})\big). \tag{3.3}$$

After looking at their construction and applying the basic uniqueness result for ordinary differential equations, one sees that

$$\begin{gathered} p^{m+1}(\,\cdot\,,x,\mathbf{w}) = p^m(\,\cdot\,,x,\mathbf{w}) \quad \text{on } \big[0,\zeta^m(x,\mathbf{w})\big) \\ \text{for all } x \text{ and } \mu_{\mathbb{R}^d}\text{-almost every } \mathbf{w} \in \mathfrak{W}(\mathbb{R}^d). \end{gathered} \tag{3.4}$$

In particular, this proves that, for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$,

$$\zeta^m(\,*\,,\mathbf{w}) \leq \zeta^{m+1}(\,*\,,\mathbf{w})\big).$$

Hence, if we set

$$\mathfrak{e}(x,\mathbf{w}) \equiv \sup_{m\geq 1} \zeta^m(x,\mathbf{w}), \tag{3.5}$$

then, for each $x \in \mathbb{R}^N$, $\mathbf{w} \rightsquigarrow \mathfrak{e}(x,\mathbf{w})$ is a $\{\bar{\mathcal{B}}_t : t \geq 0\}$-stopping time and, for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$,

$$\zeta^m(\,*\,,\mathbf{w}) \nearrow \mathfrak{e}(\,*\,,\mathbf{w}).$$

In particular, just as in the analysis given in the introduction, $x \rightsquigarrow \mathfrak{e}(x,\mathbf{w})$ will be continuous for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$.

Now set $\mathcal{P}(\widehat{\mathbb{R}^N}) = C\big([0,\infty);\widehat{\mathbb{R}^N}\big)$, and give $\mathcal{P}(\widehat{\mathbb{R}^N})$ the topology of uniform convergence on compact intervals. In view of the preceding, we know that there exists a measurable map $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto p(\,\cdot\,,x,\mathbf{w}) \in \mathcal{P}(\widehat{\mathbb{R}^N})$ such that:

$$\begin{gathered} p(t,x,\mathbf{w}) = \infty \quad \text{for every } x \in \mathbb{R}^N,\ \mathbf{w} \in \mathfrak{W}(\mathbb{R}^d), \text{ and } t \in [\mathfrak{e}(x,\mathbf{w}),\infty), \\ x \in \mathbb{R}^N \longmapsto p(\,\cdot\,,x,\mathbf{w}) \in \mathcal{P}(\widehat{\mathbb{R}^N}) \quad \text{is continuous for every } \mathbf{w} \in \mathfrak{W}(\mathbb{R}^d), \\ p(t,x,\mathbf{w}) = p^m(t,x,\mathbf{w}) \quad \text{for all } m \geq 1,\ 0 \leq t < \zeta^m(x,\mathbf{w}), \\ x \in \mathbb{R}^N, \text{ and } \mu_{\mathbb{R}^d}\text{-almost every } \mathbf{w} \in \mathfrak{W}(\mathbb{R}^d). \end{gathered} \tag{3.6}$$

We summarize these observations as a theorem.

3.7 THEOREM. *For each $T \in [0,\infty)$ and $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$, set $U(T,\mathbf{w}) = \{x \in \mathbb{R}^N : \mathfrak{e}(x,\mathbf{w}) > T\}$. Then for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$ and all T, $U(T,\mathbf{w})$ is open and*

$$p(\,\cdot\,,*,\mathbf{w}) \restriction [0,T] \times U(T,\mathbf{w}) \in C^{0,\infty}\big([0,T]\times\mathbb{R}^N;\mathbb{R}^N\big).$$

Moreover, if σ is a $\{\bar{\mathcal{B}}_t : t \ge 0\}$-stopping time, then (cf. (2.60) and (2.63))

$$\begin{gathered}\sigma(\mathbf{w}) < \mathfrak{e}(x,\mathbf{w}) \implies \\ \mathfrak{e}(x,\mathbf{w}) = \sigma(\mathbf{w}) + \mathfrak{e}\Big(p\big(\sigma(\mathbf{w}),x,\mathbf{w}\big),\delta_{\sigma(\mathbf{w})}\mathbf{w}\Big)\end{gathered}$$

and

$$\begin{gathered}t + \sigma(\mathbf{w}) < \mathfrak{e}(x,\mathbf{w}) \implies \\ p\big(t+\sigma(\mathbf{w}),x,\mathbf{w}\big) = p\Big(t, p\big(\sigma(\mathbf{w}),x,\mathbf{w}\big)\Big)\end{gathered}$$

for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$. Finally, if we extend $f \in C_c(\mathbb{R}^N;\mathbb{R})$ to $\widehat{\mathbb{R}^N}$ by taking $f(\infty) = 0$, then, for every $f \in C^2_c(\mathbb{R}^N;\mathbb{R})$,

$$f\big(p(t,x,\mathbf{w})\big) - \int_0^t \big[\mathcal{L}f\big]\big(p(\tau,x,\mathbf{w})\big)\,d\tau$$

is a $\mu_{\mathbb{R}^d}$-martingale relative to $\{\bar{\mathcal{B}}_t : t \ge 0\}$. In particular, if $\mathfrak{e}(x,\cdot\,) = \infty$ $\mu_{\mathbb{R}^d}$-almost surely for some $x \in \mathbb{R}^N$, then the $\mu_{\mathbb{R}^d}$-distribution of

$$\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto p(\,\cdot\,,x,\mathbf{w}) \restriction \big[0,\mathfrak{e}(x,\mathbf{w})\big) \in \mathcal{P}(\mathbb{R}^N)$$

is a solution $\mathbb{P}^{\mathcal{L}}_x$ to the martingale problem for $\mathcal{L}$ starting at x.

3.2.2. Splicing. In the preceding section we showed that a solution to the martingale problem for $\mathcal{L}$ starting at x exists when (cf. (3.5)) $\mathfrak{e}(x,\cdot\,) = \infty$ (a.s., $\mu_{\mathbb{R}^d}$). However, as yet, we have no way to check uniqueness. Indeed, although one suspects that any solution must coincide with (cf. the discussion preceding Theorem 3.7) $\mathbb{P}^{\mathcal{L}_m}_x$ on $\mathcal{F}_{\zeta^{U_m(x)}}$ for every $m \ge 1$ and is therefore unique, verification of this suspicion still requires some work. In fact, exactly what we are lacking is a method for slicing measures together at a stopping time.

In order to understand the reasoning which follows, it may be helpful to review what happens in the deterministic setting. Thus, suppose that X and $\tilde{X}$ are smooth vector fields which agree on a bounded open set U. Further, assume that $\tilde{X}$ is bounded everywhere, and, for each $x \in \mathbb{R}^N$, let $\bar{p}_x$ be the integral curve of $\tilde{X}$ starting at x. Finally, let $p : \big[0,\zeta^U(p)\big) \longrightarrow U$ be an integral curve of X starting at some $x_0 \in U$, and build the *spliced* path $q = p_x \underset{\zeta^U}{\otimes} \bar{p}$. so that

$$q(t) = \begin{cases} p(t) & \text{if } t \in \big[0,\zeta^U(p)\big) \\ \bar{p}_{p(\zeta^U)}\big(t-\zeta^U(p)\big) & \text{if } \zeta^U(p) \le t < \infty. \end{cases}$$

Then, q is an integral curve of $\bar{X}$ starting at x_0, and therefore $q = \bar{p}_{x_0}$, which means, in particular, that p coincides with $\bar{p}_{x_0}$ on $[0, \zeta^U(p))$.[2]

Our purpose here is to develop the analogous observation when integral curves are replaced by solutions to the martingale problem; and this will require us to learn how *splice* measures together at a stopping time. To get started, let $\mathbf{M}_1(\mathcal{P}(\mathbb{R}^N))$ denote the space of Borel probability measures on $\mathcal{P}(\mathbb{R}^N)$, and suppose that $\{\bar{\mathbb{P}}_x : x \in \mathbb{R}^N\} \subseteq \mathbf{M}_1(\mathcal{P}(\mathbb{R}^N))$ has the properties that $x \in \mathbb{R}^N \longmapsto \bar{\mathbb{P}}_x(\Gamma) \in [0,1]$ is measurable for every Borel set $\Gamma \subseteq \mathcal{P}(\mathbb{R}^N)$ and that $p(0) = x$ $\bar{\mathbb{P}}_x$-almost surely for each $x \in \mathbb{R}^N$. Next, define

$$(t, q) \in [0, \infty) \times \mathcal{P}(\mathbb{R}^N) \longmapsto \delta_q \underset{t}{\otimes} \bar{\mathbb{P}}_\cdot \in \mathbf{M}_1(\mathcal{P}(\mathbb{R}^N))$$

so that, for $0 \le \sigma_1 < \cdots < \sigma_m < t \le \tau_1 < \cdots < \tau_n$ and Borel subsets $A_1, \dots, A_m, B_1, \dots, B_n$ in $\mathbb{R}^N$:

$$\begin{aligned}\delta_q \underset{t}{\otimes} \bar{\mathbb{P}}_\cdot \Big(p(\sigma_1) \in A_1, \dots, p(\sigma_m) \in A_m, p(\tau_1) \in B_1, \dots, p(\tau_n) \in B_n\Big) \\ = \mathbf{1}_{A_1}(q(\sigma_1)) \cdots \mathbf{1}_{A_m}(q(\sigma_m)) \bar{\mathbb{P}}_{q(t)}\Big(p(\tau_1 - t) \in B_1, \dots, p(\tau_n - t) \in B_n\Big).\end{aligned}$$

Finally, given a $\mathbb{P} \in \mathbf{M}_1(\mathcal{P}(\mathbb{R}^N))$ and a stopping time $\zeta : \mathcal{P}(\mathbb{R}^N) \longrightarrow [0, \infty]$, define the *spliced measure* $\mathbb{P} \underset{\zeta}{\otimes} \bar{\mathbb{P}}_\cdot \in \mathbf{M}_1(\mathcal{P}(\mathbb{R}^N))$ so that

$$\mathbb{P} \underset{\zeta}{\otimes} \bar{\mathbb{P}}_\cdot(A) = \int_{\{\zeta < \infty\}} \delta_q \underset{\zeta(q)}{\otimes} \bar{\mathbb{P}}_\cdot(A) \mathbb{P}(dq) + \mathbb{P}(A \cap \{\zeta = \infty\}). \tag{3.8}$$

In order to bring out the analogy between this construction and the one for paths, it is useful to provide another description of the measure $\mathbb{P} \underset{\zeta}{\otimes} \bar{\mathbb{P}}_\cdot$. To this end, define the *shift map* $\Sigma_t : \mathcal{P}(\mathbb{R}^N) \longrightarrow \mathcal{P}(\mathbb{R}^N)$ so that $[\Sigma_t(p)](\,\cdot\,) = p(\,\cdot\, + t)$. Then $\mathbb{P} \underset{\zeta}{\otimes} \bar{\mathbb{P}}_\cdot$ is completely determined by the properties that its restriction to $\mathcal{F}_\zeta$ is $\mathbb{P}$ and, given $\mathcal{F}_\zeta$, the conditional distribution of shifted path

$$p \in \{\zeta < \infty\} \longmapsto \Sigma_{\zeta(p)} p \in \mathcal{P}(\mathbb{R}^N) \quad \text{under } \mathbb{P} \underset{\zeta}{\otimes} \bar{\mathbb{P}}_\cdot$$

is $\bar{\mathbb{P}}_{p(\zeta)}$. Using this interpretation, one can easily verify (cf. Theorem 1.2.10 in [42]) the following important technical fact about $\mathbb{P} \underset{\zeta}{\otimes} \bar{\mathbb{P}}_\cdot$.

3.9 LEMMA. *Let $M : [0, \infty) \times \mathcal{P}(\mathbb{R}^N) \longrightarrow \mathbb{R}$ be a continuous,*[3] *$\{\mathcal{F}_t : t \ge 0\}$-adapted function whose restriction to $[0, T] \times \mathcal{P}(\mathbb{R}^N)$ is bounded for every*

---

[2] If one thinks about it, one realizes that this is the observation which allowed us to assert in the introductory discussion that p^{m+1} coincides with p^m on $[0, \zeta^m)$.

[3] The terminology here is simultaneously confusing and conventional. The *continuity* refers to $t \rightsquigarrow M(t, \mathrm{p})$ for each $\mathrm{p} \in \mathcal{P}(\mathbb{R}^N)$, it does not mean that $\mathrm{p} \rightsquigarrow M(t, \mathrm{p})$ is continuous.

$T \in [0,\infty)$. Further, assume that $\big(M(t\wedge\zeta),\mathcal{F}_t,Q\big)$ is a martingale and that, for each $q \in \{\zeta<\infty\}$,

$$\Big(M(t)-M\big(t\wedge\zeta(q)\big),\mathcal{F}_t,\delta_q \underset{\zeta(q)}{\otimes} \bar{\mathbb{P}}.\Big)$$

is a martingale. Then,

$$\Big(M(t),\mathcal{F}_t,\mathbb{P}\underset{\zeta}{\otimes}\bar{\mathbb{P}}.\Big)$$

is also a martingale.

3.2.3. Localizing the Martingale Problem. We are now in a position[4] to carry out the localization procedure alluded to at the beginning of the preceding section.

3.10 THEOREM. *Let $\{X_k : 0 \le k \le d\}$ and $\{\tilde{X}_k : 0 \le k \le d\}$ be two collections of smooth vector fields on $\mathbb{R}^N$, and define the operators $\mathcal{L}$ and $\tilde{\mathcal{L}}$ accordingly, as in (2.3). Next, assume that the $\tilde{X}_k$'s have bounded derivatives of every order, and let $\{\mathbb{P}_x^{\tilde{\mathcal{L}}} : x \in \mathbb{R}^N\}$ be the associated family of solutions to the martingale problem for $\tilde{\mathcal{L}}$, as guaranteed by Theorems 2.40 and 2.68. Finally, let U be a non-empty, open subset of $\mathbb{R}^N$, and assume that $X_k \restriction U = \tilde{X}_k \restriction U$ for each $0 \le k \le d$. Then, for any solution $\mathbb{P}$ to the martingale problem for $\mathcal{L}$ starting at some $x_0 \in U$,*

$$\mathbb{P}_{x_0}^{\tilde{\mathcal{L}}} = \mathbb{P}\underset{\zeta^U}{\otimes}\mathbb{P}_{\cdot}^{\tilde{\mathcal{L}}}\,.$$

In particular,

$$\mathbb{P}\restriction\mathcal{F}_{\zeta^U} = \mathbb{P}_{x_0}^{\tilde{\mathcal{L}}}\restriction\mathcal{F}_{\zeta^U}.$$

PROOF: There is hardly anything to do. Namely, given $f \in C_c^\infty(\mathbb{R}^N;\mathbb{R})$, set

$$M(t,p) = f\big(p(t)\big) - \int_0^t \big[\tilde{\mathcal{L}}f\big]\big(p(\tau)\big)\,d\tau.$$

Because $\mathcal{L}f \restriction U = \tilde{\mathcal{L}}f \restriction U$, Doob's Stopping Time Theorem assures us that $M(t\wedge\zeta^U)$ is a $\mathbb{P}_{x_0}$-martingale. At the same time, his theorem says that

$$M(t)-M\big(t\wedge\zeta^U(q)\big) \text{ will be a } \delta_q\underset{\zeta^U}{\otimes}\mathbb{P}_{\cdot}^{\tilde{\mathcal{L}}}\text{-martingale for every } q\in\mathcal{P}(\mathbb{R}^N).$$

Thus, by Lemma 3.9, $M(t)$ is a $\mathbb{P}_{x_0}\underset{\zeta^U}{\otimes}\mathbb{P}_{\cdot}^{\tilde{\mathcal{L}}}$-martingale. Since this is true for every $f \in C_c^\infty(\mathbb{R}^N;\mathbb{R})$, Theorem 2.68 now gives to the desired conclusion. □

As an immediate consequence of the preceding, we have the following statement of uniqueness.

[4] Actually, we have been indulging in a bit of overkill here: there are more elementary ways to prove the uniqueness result given below. However, the localization machinery developed here has advantages not shared by other approaches.

3.11 COROLLARY. *Let $(X_0, \dots, X_d)$ be a family of smooth vector fields on $\mathbb{R}^N$, and define $\mathcal{L}$ accordingly, as in (2.3). Then, for each $x \in \mathbb{R}^N$, there is at most one solution $\mathbb{P}_x^{\mathcal{L}}$ to the martingale problem for $\mathcal{L}$ starting at x. Moreover, if (cf. (3.5))*

$$\mathfrak{e}(x, \cdot\,) = \infty \quad (\text{a.s.}, \mu_{\mathbb{R}^d}), \tag{3.12}$$

then the one and only solution to the martingale problem for $\mathcal{L}$ starting at x is the measure $\mathbb{P}_x^{\mathcal{L}}$ described in Theorem 3.7.

3.2.4. A Non-explosion Criterion. For obvious reasons, the solution to the martingale problem for $\mathcal{L}$ is said to *explode* when (3.12) fails, but, just as is the case with integral curves, explosion need not be seen as a disaster. For instance, one can talk about solutions *up to the time of explosion*, and there is no problem applying Theorem 3.10 to see that, up to the explosion time, the solution will be unique. In addition, it is often useful to consider continuations of the solution beyond the explosion time. However, such continuations are will not be considered here. Instead, we will provide the following, rather crude, criterion for non-explosion.

3.13 THEOREM. *Suppose that there exists a $u \in C^2(\mathbb{R}^N; [0,\infty))$ with the properties that*

$$\lambda u - \mathcal{L}u \geq 0 \text{ for some } \lambda > 0 \quad \text{and} \quad \lim_{|x|\to\infty} u(x) = \infty.$$

Then, for every $x \in \mathbb{R}^N$, (3.12) holds and so the martingale problem for $\mathcal{L}$ has precisely one solution $\mathbb{P}_x^{\mathcal{L}}$ starting at x.

PROOF: Given $m \geq 1$, choose $f_m \in C_b^2([0,\infty) \times \mathbb{R}^N; [0,\infty))$ so that $f_m(t,x) = e^{-\lambda t} u(x)$ when $x \in U_m$ (cf. § 3.2.1). Then, because

$$\frac{\partial f_m}{\partial t} + \mathcal{L}_m f_m \leq 0 \quad \text{on } [0,\infty) \times U_m,$$

Theorem 2.40 combined with Doob's Stopping Time Theorem shows that

$$\mathbb{E}^{\mathbb{P}_x^{\mathcal{L}_m}}\left[e^{-\lambda \zeta^{U_m}} u\big(p(\zeta^{U_m}), \zeta^{U_m} < \infty\right] \leq u(x).$$

Thus, if $\alpha_m \equiv \min_{x \notin U_m} u(x)$, then

$$\alpha_m \mathbb{E}^{\mathbb{P}_{x_0}^{\mathcal{L}_m}}\left[e^{-\lambda \zeta^{U_m}}, \zeta^{U_m} < \infty\right] \leq u(x_0).$$

Since $\alpha_m \to \infty$ as $m \to \infty$, this means

$$\mathbb{P}_{x_0}^{\mathcal{L}_m}\left(\zeta^{U_m} \leq T\right) \leq e^{\lambda T} \mathbb{E}^{\mathbb{P}_x^{\mathcal{L}_m}}\left[e^{-\lambda \zeta^{U_m}}, \zeta^{U_m} < \infty\right] \longrightarrow 0$$

as $m \to \infty$ for every $T \in [0,\infty)$. □

3.14 COROLLARY. *If the coefficients of the X_k's have bounded first order derivatives, then the martingale problem for $\mathcal{L}$ has precisely one solution starting each $x \in \mathbb{R}^N$.*

PROOF: Take $u(x) = 1 + |x|^2_{\mathbb{R}^N}$. From the hypotheses, we know that there is a $C < \infty$ such that $[\mathcal{L}u](x) \leq C(1 + |x|^2_{\mathbb{R}^N})$. Hence, we can take $\lambda = C$. □

3.2.5. Well-posed Martingale Problems. When the martingale problem for some operator has precisely precisely one solution starting at each $x \in \mathbb{R}^N$, we will say that that martingale problem is *well-posed.* As was shown in [42] (especially Chapter 6), abstract considerations lead, without reference to any particular setting, to the conclusion that well-posed martingale problems possess a great many desirable properties. However, it is possible to check most of these properties by hand when dealing with the setting under consideration here. In fact, the whole of the next statement contains nothing which is not already implicit in the conjunction of Theorems 3.7 with 3.10.

3.15 THEOREM. *Let $\mathcal{L}$ be the operator associated with smooth vector fields $(X_0, \dots, X_d)$ on $\mathbb{R}^N$. Then the martingale problem for $\mathcal{L}$ is well-posed if and only if (3.12) holds for every $x \in \mathbb{R}^N$. Moreover, if the martingale problem for $\mathcal{L}$ is well-posed and $\{\mathbb{P}^{\mathcal{L}}_x : x \in \mathbb{R}^N\}$ is the corresponding family of solutions, then, $x \in \mathbb{R}^N \longmapsto \mathbb{P}^{\mathcal{L}}_x \in \mathbf{M}_1(\mathcal{P}(\mathbb{R}^N))$ is continuous*[5] *and, for each $x \in \mathbb{R}^N$ and $\{\mathcal{F}_t : t \geq 0\}$-stopping time ζ,*

$$\mathbb{P}^{\mathcal{L}}_x = \mathbb{P}^{\mathcal{L}}_x \underset{\zeta}{\otimes} \mathbb{P}^{\mathcal{L}}_{\cdot}. \tag{3.16}$$

The equation (3.16) is a particularly powerful formulation of the *Markov property.*[6] In that it says that the Markov property holds at all stopping times, it is saying, in the jargon of probability theory, that $\{\mathbb{P}^{\mathcal{L}}_x : x \in \mathbb{R}^N\}$ is a *strong Markov family.*

3.3 A Polygonal Approximation Scheme

When confronted with the problem of rationalizing (2.13) for generic $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$, there is a plethora of schemes which come to mind; and among the foremost of these is the idea of first replacing $\mathbf{w}$ by a mollified version of itself, then solving (2.13) for the mollified $\mathbf{w}$, and finally turning the mollification off. The purpose of this section is to show that the simplest such scheme does indeed work. Namely, given $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$, define the polygonal approximation $\mathbf{w}_n$ for $n \in \mathbb{N}$ so that (cf. §2.3.1)

$$\mathbf{w}_n(t) = 2^n\Big(\big(T_{m+1,n} - t\big)\mathbf{w}(T_{m,n}) + \big(t - T_{m,n}\big)\mathbf{w}(T_{m+1,n})\Big) \\ \text{for } m \in \mathbb{N} \text{ and } t \in [T_{m,n}, T_{m+1,n}]. \tag{3.17}$$

[5] Continuity here refers to what probabilists call the weak topology for the space of probability measures on $\mathcal{P}(\mathbb{R}^N)$. What it means is that continuity holds when $x \rightsquigarrow \mathbb{P}^{\mathcal{L}}_x$ is tested against bounded continuous functions on $\mathcal{P}(\mathbb{R}^N)$.

[6] In the abstract, a dynamical system has the Markov property if it evolves in such a way that a question about its future behavior can be answered on the basis of information about its position at the instant when the question is asked.

We will show that, at least up to the time of explosion, one can approximate $p(\,\cdot\,, x, \mathbf{w})$ by $p(\,\cdot\,, x, \mathbf{w}_n)$, where the latter is the solution to (2.13) with $\mathbf{w}_n$ in place of $\mathbf{w}$.

3.3.1. The Bounded Case. We begin in the setting of §2.3, where each of the X_k's has bounded derivatives of all orders. Given $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$, the basic uniqueness theorem for ordinary differential equations tells us that (cf. (2.17))

$$p_n\big(T_{m,n}, x, \mathbf{w}\big) = p\big(T_{m,n}, x, \mathbf{w}_n\big) \quad \text{for all } x \in \mathbb{R}^N \text{ and } m,\, n \in \mathbb{N}.$$

At the same time, there is a $C < \infty$ such that, for any $T_{m,n} \le t \le T_{m+1,n}$,

$$\big|p_n(t, x, \mathbf{w}) - p_n\big(T_{m,n}, x, \mathbf{w}\big)\big|_{\mathbb{R}^N} \le C\Big((t - T_{m,n}) + \big|\mathbf{w}(t) - \mathbf{w}(T_{m,n})\big|_{\mathbb{R}^d}\Big)$$

and

$$\big|p(t, x, \mathbf{w}_n) - p\big(T_{m,n}, x, \mathbf{w}_n\big)\big|_{\mathbb{R}^N} \le C\Big(1 + 2^n\big|\Delta_{m,n}\mathbf{w}\big|_{\mathbb{R}^d}\Big)\big(t - T_{m,n}\big),$$

where (cf. § 1.1.1) $\Delta_{m,n}\mathbf{w} \equiv \mathbf{w}(T_{m+1,n}) - \mathbf{w}(T_{m,n})$. Hence, for any $T \in [0, \infty)$,

$$\begin{aligned}\sup_{t\in[0,T]}\ \sup_{x\in\mathbb{R}^N} &\big|p_n(t, x, \mathbf{w}) - p(t, x, \mathbf{w}_n)\big|_{\mathbb{R}^N} \\ &\le C\left(1 + \sup_{0\le s<t\le 1+T} \frac{\big|\mathbf{w}(t) - \mathbf{w}(s)\big|_{\mathbb{R}^d}}{(t - s)^{\frac{1}{4}}}\right) 2^{-\frac{n}{4}}.\end{aligned}$$

By combining the preceding, the first part of Theorem 2.58 (with $\beta = \mathbf{0}$), and (1.6), we now see that, for all $T \in [0, \infty)$ and $R \in (0, \infty)$:

$$\begin{gathered}\lim_{n\to\infty}\ \sup_{t\in[0,T]}\ \sup_{|x|\le R} \big|p(t, x, \mathbf{w}_n) - p(t, x, \mathbf{w})\big|_{\mathbb{R}^N} \longrightarrow 0 \\ \text{both } \mu_{\mathbb{R}^d}\text{-almost surely and in } L^2(\mu_{\mathbb{R}^d}; \mathbb{R}).\end{gathered}$$

Clearly, the preceding shows that an polygonal approximation scheme works, at least to zeroth order. Moreover, to show that it works to all orders, we can simply repeat the argument in § 2.4, only this time for $p(\,\cdot\,, *\,, \mathbf{w}_n)$ in place of $p_n(\,\cdot\,, *\,, \mathbf{w})$, after which we can say that

$$\begin{gathered}\lim_{n\to\infty}\ \sup_{t\in[0,T]}\ \sup_{|x|_{\mathbb{R}^N}\le R} \big|\partial_x^\beta p(t, *, \mathbf{w}_n) - \partial_x^\beta p(t, *, \mathbf{w})\big|_{\mathbb{R}^N} \longrightarrow 0 \\ \text{both } \mu_{\mathbb{R}^d}\text{-almost surely and in } L^2(\mu_{\mathbb{R}^d}; \mathbb{R}) \\ \text{for every } T \in [0, \infty),\ R \in (0, \infty), \text{ and } \beta \in \mathbb{N}^N.\end{gathered} \tag{3.18}$$

3.3.2. The General Case. To handle general smooth X_k's, we need only use the machinery developed in § 3.3 together with (3.18). That is, choose an increasing exhaustion $\{U_m\}_1^\infty$ of $\mathbb{R}^N$ by bounded open sets, and define

$\{(X_{0,m},\dots,X_{d,m}) : m \geq 1\}$ accordingly, as in §3.2.1. Next, for $n \in \mathbb{N}$ and $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$, let $p^m(\,\cdot\,,x,\mathbf{w}_n)$ be the solution to (2.13) when $\mathbf{w}$ and X_k are replaced, respectively, by $\mathbf{w}_n$ and $X_{k,m}$. Then (cf. the introduction to §3.3), for each n and $\mathbf{w}$, there is a unique continuous map

$$x \in \mathbb{R}^N \longmapsto p(\,\cdot\,,x,\mathbf{w}_n) \in \mathcal{P}(\widehat{\mathbb{R}^N})$$

such that

$$p(t,x,\mathbf{w}_n) = \begin{cases} p^m(t,x,\mathbf{w}_n) & \text{if } \zeta^m(x,\mathbf{w}) > t \\ \infty & \text{if } \mathfrak{e}(x,\mathbf{w}_n) \leq t < \infty, \end{cases}$$

where (cf. (3.3))

$$\mathfrak{e}(x,\mathbf{w}_n) \equiv \lim_{m\to\infty} \zeta^m(x,\mathbf{w}_n).$$

By applying the results in §3.2 and §3.3.1, we see first that (cf. (3.5))

$$\mathfrak{e}(\,\cdot\,,\mathbf{w}_n) \longrightarrow \mathfrak{e}(\,\cdot\,,\mathbf{w}) \quad \text{uniformly on compacts for } \mu_{\mathbb{R}^d}\text{-almost every } \mathbf{w},$$

and then, by (3.18), that

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \sup_{x\in U_R(T,\mathbf{w})} \left|\partial_x^\beta p(t,x,\mathbf{w}_n) - \partial_x^\beta p(t,x,,\mathbf{w})\right|_{\mathbb{R}^N} \longrightarrow 0 \tag{3.19}$$

$$\mu_{\mathbb{R}^d}\text{-almost surely for every } T \in [0,\infty),\ R \in (0,\infty), \text{ and } \beta \in \mathbb{N}^N,$$

where (cf. Theorem 3.7) $U_R(T,\mathbf{w}) \equiv \{x : |x| \leq R \;\&\; \mathfrak{e}(x,\mathbf{w}) > T\}$.

3.4 Subordination

A familiar trick when confronted with the problem of actually solving an ordinary differential equation is to look for and exploit some sort of inherent "lower triangularity" possessed by the equation under consideration. More explicitly, suppose X is a smooth complete vector field on $\mathbb{R}^N$ with the property that, for some $1 \leq N' < N$, the action of X on functions $f \in C^1(\mathbb{R}^N;\mathbb{R})$ takes the form

$$X_x f = X'_{x'} f(\,\cdot\,,x'') + \left(X''(x')\right)_{x''} f(x',\,\cdot\,) \quad \text{if } x = (x',x'') \in \mathbb{R}^{N'} \times \mathbb{R}^{N''},$$

where $N'' = N - N'$, X' is a smooth vector field on $\mathbb{R}^{N'}$, and $x' \rightsquigarrow X''(x')$ is a smooth map from $\mathbb{R}^{N'}$ into smooth vector fields on $\mathbb{R}^{N''}$. Then the problem of integrating X can be broken into two steps. Namely, the integral curve p of X starting at a point $x = (x',x'')$ can be expressed as $t \in \mathbb{R} \longmapsto (p'(t),p''(t)) \in \mathbb{R}^{N'} \times \mathbb{R}^{N''}$, where p' is the integral curve of X' starting at x' and p'' is determined by the *time-dependent* ordinary differential equation

$$\frac{d}{dt}p''(t) = \Big(X''\big(p'(t)\big)\Big)_{p''(t)} \quad \text{with } p''(0) = x''.$$

The purpose of this section is to examine the analogous technique in the stochastic setting.

3.4.1. Time Dependent Vector Fields. As the preceding introductory remarks make clear, even when one starts with time-independent vector fields, the technique for which we are looking is going to force us to consider time-dependent vector fields. Thus, we had better begin by checking that the construction scheme introduced in §2.3 survives the transition to the time-dependent context.

Actually, as long as the time-dependence is smooth, there is a trick which allows one to reduce the time-dependent case to the time-independent case. Namely, if, for each $0 \le k \le d$, $t \rightsquigarrow X_k(t)$ is a smooth function from $\mathbb{R}$ into smooth vector fields on $\mathbb{R}^N$, define the vector fields $(\hat{X}_0, \dots, \hat{X}_d)$ on $\mathbb{R} \times \mathbb{R}^N$ so that

$$\begin{aligned}\left(\hat{X}_0\right)_{\hat{x}} f &= \partial_{x^0} f + \left(X_0(x^0)\right)_x f(x^0, \,\cdot\,)\\ \left(\hat{X}_k\right)_{\hat{x}} f(x^0, \,\cdot\,) &= \left(X_k(x^0)\right)_x f(x^0, \,\cdot\,) \quad \text{when } 1 \le k \le d\end{aligned}$$

for $\hat{x} = (x^0, x) \in \mathbb{R} \times \mathbb{R}^N$. Assuming that the $\hat{X}_k$'s have bounded derivatives of all orders, we can use the results in §2.3 to construct the map

$$\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto \hat{p}(\,\cdot\,, \hat{x}, \mathbf{w}) \in \mathcal{P}(\mathbb{R} \times \mathbb{R}^N)$$

whose $\mu_{\mathbb{R}^d}$-distribution solves the martingale problem for

$$\hat{\mathcal{L}} = \hat{X}_0 + \frac{1}{2} \sum_{k=1}^{d} \hat{X}_k^2$$

starting at $\hat{x}$. In particular, if

$$\mathcal{L}_t \equiv X_0(t) + \frac{1}{2} \sum_{k=1}^{d} X_k(t)^2 \tag{3.20}$$

and we write $\hat{p}(t, (0, x), \mathbf{w}) = (t, p(t, x, \mathbf{w}))$, then, $p(0, x, \mathbf{w}) = x$ and, for every $f \in C_{\mathrm{b}}^2(\mathbb{R}^N; \mathbb{R})$,

$$f\left(p(t, x, \mathbf{w})\right) - \int_0^t \left[\mathcal{L}_\tau f\right]\left(p(\tau, x, \mathbf{w})\right) d\tau$$

is a $\mu_{\mathbb{R}^d}$-martingale relative to $\{\bar{\mathcal{B}}_t : t \ge 0\}$. In the classical literature, this trick is known as reduction of time-inhomogeneous processes to time-homogeneous ones by consideration of the *time-space* process.

Unfortunately, even when we start with smooth vector fields, we get less than smooth time-dependence after composition with stochastic paths. Thus, the preceding trick is not sufficient to our needs, and we have to re-examine the contents of §2.3 to make sure that all is not lost.

The setting in which we will be working is the following. For each $0 \le k \le d$, $t \rightsquigarrow X_k(t)$ satisfies

$$(3.21) \quad \sup_{(s,x)\in[0,\infty)\times\mathbb{R}^N} \left(\left|\partial_x^\alpha X_k(s)\right|_{\mathbb{R}^N} \vee \sup_{0<t-s\le 1} \frac{\left|\partial_x^\alpha X_k(t) - \partial_x^\alpha X_k(s)\right|_{\mathbb{R}^N}}{(t-s)^{\frac{1}{4}}} \right) < \infty$$
$$\text{for each } \alpha \in \mathbb{N}^N.$$

Next, given $\Xi = (\xi^0, \boldsymbol{\xi}) \in [0,\infty) \times \mathbb{R}^d$, set

$$X_\Xi(t) = \sum_{k=0}^d \xi^k X_k(t),$$

and use the elementary theory of ordinary differential equations to see that, for each $(s,x) \in [0,\infty) \times \mathbb{R}^N$, there is a smooth map $\Xi \in [0,\infty) \times \mathbb{R}^d \longmapsto E(\Xi, (s,x)) \in \mathbb{R}^N$ so that

$$E\big(\Xi,(s,x)\big) = E\big(1,\Xi,(s,x)\big) \quad \text{where}$$
$$\frac{d}{dt} E\big(t,\Xi,(s,x)\big) = X_\Xi(s+\xi^0 t)_{E(t,\Xi,(s,x))} \quad \text{with } E\big(0,\Xi,(s,x)\big) = x.$$

In fact, for each $S \in (0,\infty)$ and $(\alpha,\beta) \in \mathbb{N}^{d+1} \times \mathbb{N}^N$,

$$(3.22) \quad \sup_{(s,x)\in[0,S]\times\mathbb{R}^N} \ \sup_{|\Xi|_{\mathbb{R}^{d+1}}\le S} \left|\partial_\Xi^\alpha \partial_x^\beta \bar{E}\big(\Xi,(s,x)\big)\right|_{\mathbb{R}^N} < \infty$$
$$\text{where } \bar{E}\big(\Xi,(s,x)\big) \equiv E\big(\Xi,(s,x)\big) - x.$$

Next (cf. (2.16)), for each $n \in \mathbb{N}$, $x \in \mathbb{R}^N$, and $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$, define $t \rightsquigarrow p_n(t,x,\mathbf{w})$ so that

$$(3.23) \quad \begin{gathered} p_n(0,x,\mathbf{w}) = x \text{ and, for } t \in [T_{m,n}, T_{m+1,n}], \\ p_n(t,x,\mathbf{w}) = E\Big(\Xi_{(m,n)}(t,\mathbf{w}), \big(T_{m,n}, p_n(T_{m,n},x,\mathbf{w})\big)\Big). \end{gathered}$$

What we would like to prove is that the obvious analog of Theorem 2.40 continues to hold. In particular, we want to know that there exists a measurable $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto p(\,\cdot\,,x,\mathbf{w}) \in \mathcal{P}(\mathbb{R}^N)$ such that

$$(3.24) \quad \begin{gathered} \lim_{n\to\infty} \sup_{t\in[0,T]} \left|p_n(t,x,\mathbf{w}) - p(t,x,\mathbf{w})\right|_{\mathbb{R}^N} = 0 \\ \mu_{\mathbb{R}^d}\text{-almost surely and in } L^2(\mu_{\mathbb{R}^d};\mathbb{R}) \end{gathered}$$

for each $(T,x) \in [0,\infty) \times \mathbb{R}^N$. If such a $\mathbf{w} \rightsquigarrow p(\,\cdot\,,x,\mathbf{w})$ exists, then a repetition of the argument given to prove Theorem 2.40 shows that, for each $f \in C_{\mathrm{b}}^{1,2}\big([0,\infty) \times \mathbb{R}^N;\mathbb{R}\big)$, (cf. (3.20))

$$(3.25) \quad \begin{gathered} f\big(t,p(t,x,\mathbf{w})\big) - \int_0^t \left(\frac{\partial f}{\partial \tau} + \mathcal{L}_\tau f\right)\big(\tau, p(\tau,x,\mathbf{w})\big)\,d\tau \\ \text{is a } \mu_{\mathbb{R}^d}\text{-martingale relative to } \{\bar{\mathcal{B}}_t : t \ge 0\}. \end{gathered}$$

After a careful look at the arguments which led to Theorem 2.40, one realizes that the only serious differences between what was done there and what is needed here all stem from the loss of (2.20). Thus, we should begin by finding the appropriate substitute for (2.20) in the present setting. To begin with, observe that, for any test function φ,

$$\begin{aligned}\varphi\big(E(\Xi,(s,x))\big) =&\varphi(x) + \int_0^1 X_{\Xi}(s+\tau\xi^0)_x\varphi\,d\tau\\ &+ \iint\limits_{0\le\tau_2\le\tau_1\le1} X_{\Xi}(s+\tau_2\xi^0)_{E(\tau_2,(s,x))}X_{\Xi}(s+\tau_1\xi^0)\varphi\,d\tau_1 d\tau_2.\end{aligned}$$

Hence, if $\bar{E}\big(\Xi,(s,x)\big) \equiv E\big(\Xi,(s,x)\big) - x$, then

$$\begin{aligned}\bar{E}\big(\Xi,(s,x)\big) = &\int_0^1 X_{\Xi}(s+\tau\xi^0)\,d\tau\\ &+ \iint\limits_{0\le\tau_2\le\tau_1\le1} X_{\Xi}(s+\tau_2\xi^0)_x X_{\Xi}(s+\tau_1\xi^0)\,d\tau_1 d\tau_2 + \mathcal{O}\big(|\Xi|_{\mathbb{R}^{d+1}}|^3\big).\end{aligned}$$

If we now use this to compare $E\big(\Xi',(s+\xi^0,E(\Xi,(s,x))\big)$ with $E\big(\Xi+\Xi',(s,x)\big)$, we find that

$$\begin{aligned}&E\big(\Xi',(s+\xi^0,E(\Xi,(s,x))\big) - E\big(\Xi+\Xi',(s,x)\big)\\ &\quad= Y_{\Xi,\Xi'}(s)_x + \iint\limits_{0\le\tau_2\le\tau_1\le1}\big[X_{\Xi'}(s+\tau_1\eta),X_{\Xi}(s+\tau_2\eta)\big]_x\,d\tau_1 d\tau_2\\ &\qquad\qquad + R_3\big(\Xi,\Xi',(s,x)\big),\end{aligned} \tag{3.26}$$

where $\eta \equiv \xi^0 + (\xi')^0$,

$$\begin{aligned}Y_{\Xi,\Xi'}(s) = &\sum_{k=1}^d \xi^k \int_0^1 \Big(X_k\big(s+\tau\xi^0\big) - X_k\big(s+\tau\eta\big)\Big)\,d\tau\\ &+ \sum_{k=1}^d (\xi')^k \int_0^1 \Big(X_k\big(s+\xi^0+\tau(\xi')^0\big) - X_k\big(s+\tau\eta\big)\Big)\,d\tau,\end{aligned}$$

and there is (cf. (3.21)) a $C<\infty$ such that

$$\big|R_3\big(\Xi,\Xi',(s,x)\big)\big|_{\mathbb{R}^N} \le C\big(|\Xi|^{\frac{9}{4}}_{\mathbb{R}^{d+1}} + |\Xi'|^{\frac{9}{4}}_{\mathbb{R}^{d+1}} + \eta^{\frac{5}{4}}\big).$$

(3.26) is used in the estimation of (cf. (2.22)) $\bar{\Delta}_{(m,n)}$. Namely, when we segregate out the martingale part $\tilde{\mathbf{M}}_{(m,n)}$ of $\tilde{\Delta}_{(m,n)}$ as we did in (2.33), we now need to take into account the contribution of terms coming from $Y_{\Xi,\Xi'}$. Thus, the right hand side of (2.34) gets replaced by

$$\begin{aligned}
&\sum_{k=1}^{d} \xi^k_{(2m,n+1)}(t,\mathbf{w}) \int_0^1 \Big(X_k\big(T_{m,n} + \tau \xi^0_{(2m,n+1)}(t)\big) \\
&\qquad\qquad - X_k\big(T_{m,n} + \tau \xi^0_{(m,n)}(t)\big)\Big)_{p_n(T_{m,n},x,\mathbf{w})} d\tau \\
&\quad + \sum_{k=1}^{d} \xi^k_{(2m+1,n+1)}(t,\mathbf{w}) \int_0^1 \Big(X_k\big(T_{m,n} + \xi^0_{(2m,n+1)}(t) + \tau \xi^0_{(2m+1,n+1)}(t)\big) \\
&\qquad\qquad - X_k\big(T_{m,n} + \tau \xi^0_{(m,n)}(t)\big)\Big)_{p_n(T_{m,n},x,\mathbf{w})} d\tau \\
&\quad + \sum_{1\le k\ne \ell\le d} \xi^k_{(2m+1,n+1)}(t,\mathbf{w}) \xi^\ell_{(2m,n+1)}(t,\mathbf{w}) \\
&\qquad \times \iint_{0\le \tau_2\le \tau_1\le 1} \big[X_k(T_{m,n}+\tau_1\eta), X_\ell(s+\tau_2\eta)\big]_{p_n(T_{m,n},x,\mathbf{w})} d\tau_1 d\tau_2.
\end{aligned}$$

Just as in the derivation of (2.36), one can go from the expression above for $\tilde{\mathbf{M}}_{(m,n)}$ to the estimate

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\sup_{t\in[0,T]} \big|\tilde{\mathbf{M}}_n(t,x,\mathbf{w})\big|^2_{\mathbb{R}^N}\right] \le C \sum_{m<2^nT} \xi^0(T)^{\frac{3}{2}},$$

where

$$\tilde{\mathbf{M}}_n(t,x,\mathbf{w}) \equiv \sum_{m=0}^{\infty} \tilde{\mathbf{M}}_{(m,n)}(t,x,\mathbf{w}).$$

The preceding estimate dispenses with the one substantive new issue introduced by the consideration of time-dependent vector fields. In other words, by recycling the techniques used in §§ 2.3, 2.4, and 2.5, one can now prove the following generalizations of Theorems 2.40, 2.58, and 2.68. In this connection, we say that a probability measure on $\mathcal{P}(\mathbb{R}^N)$ solves the martingale problem for (cf. (3.20)) $t\rightsquigarrow\mathcal{L}_t$ starting at $x\in\mathbb{R}^N$ if

$$\mathbb{P}\big(p(0)=x\big)=1 \quad\text{and}\quad f\big(p(t)\big) - \int_0^t \big[\mathcal{L}_\tau f\big]\big(p(\tau)\big)\,d\tau$$

is a $\mathbb{P}$-martingale relative $\{\mathcal{F}_t : t\ge 0\}$ for every $f\in C_c^\infty(\mathbb{R}^N;\mathbb{R})$.

3.27 THEOREM. *Let $t\rightsquigarrow\big(X_0(t),\dots,X_d(t)\big)$ be a mapping from $[0,\infty)$ into a family of smooth vector fields on $\mathbb{R}^N$, and assume that (3.21) holds. If, for each $x\in\mathbb{R}^N$, $\mathbf{w}\in\mathfrak{W}(\mathbb{R}^d)\longmapsto\{p_n(\,\cdot\,,x,\mathbf{w}) : n\ge 0\}$ is defined by (3.23), then there exists a measurable*

$$\mathbf{w}\in\mathfrak{W}(\mathbb{R}^d)\longmapsto p(\,\cdot\,,*,\mathbf{w})\in C^{0,\infty}\big([0,\infty)\times\mathbb{R}^N;\mathbb{R}^N\big)$$

such that, for each $T \in [0,\infty)$, $R \in (0,\infty)$, *and* $\beta \in \mathbb{N}^N$,

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \sup_{|x|_{\mathbb{R}^N}\le R} \left|\partial_x^\beta p_n(t,x,\mathbf{w}) - \partial_x^\beta p(t,x,\mathbf{w})\right|_{\mathbb{R}^N} = 0$$

both $\mu_{\mathbb{R}^d}$*-almost surely and in* $L^2(\mu_{\mathbb{R}^d};\mathbb{R})$. *Moreover, if* $t\rightsquigarrow\mathcal{L}_t$ *is given by (3.20), then, for each* $x \in \mathbb{R}^N$, *the* $\mu_{\mathbb{R}^d}$*-distribution of* $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto p(\,\cdot\,,x,\mathbf{w}) \in \mathcal{P}(\mathbb{R}^N)$ *is the one and only probability measure* $\mathbb{P}$ *on* $\mathcal{P}(\mathbb{R}^N)$ *which solves the martingale problem for* $t\rightsquigarrow\mathcal{L}_t$ *starting at* x.

3.4.2. Subordination for Diffusions. We now have all the machinery needed to carry out the program alluded to in the introduction to this section.

Let $(X_0,\dots,X_d)$ be a family of smooth vector fields on $\mathbb{R}^N$. Next, assume that there exists an $1 \le N' < N$ and a $1 \le d' \le d$ such that, for $x = (x',x'') \in \mathbb{R}^{N'} \times \mathbb{R}^{N''}$:

$$(3.28)\quad (X_k)_x f = \begin{cases} (X_0')_{x'} f(\,\cdot\,,x'') + \big(X_0''(x')\big)_{x''} f(x',\,\cdot\,) & \text{if } k=0 \\ (X_k')_{x'} f(\,\cdot\,,x'') & \text{if } 1\le k\le d' \\ \big(X_k''(x')\big)_{x''} f(x',\,\cdot\,) & \text{if } d'+1\le k\le d \end{cases}$$

where $N'' \equiv N - N'$, $(X_0',\dots,X_{d'}')$ are smooth vector fields on $\mathbb{R}^{N'}$, and, for $k \in \{0, d'+1,\dots,d\}$, $x'\rightsquigarrow X''(x')$ is a smooth map from $\mathbb{R}^{N'}$ into smooth vector fields on $\mathbb{R}^{N''}$.

3.29 THEOREM. *Assume that* $(X_0,\dots,X_d)$ *are vector fields which satisfy (3.28). Further, let* $x = (x',x'')$ *be given, and assume that the martingale problem for* $\mathcal{L}$ *starting at* x *has a solution (i.e. (3.12)) holds). Now, set*

$$\mathcal{L}' = X_0' + \frac{1}{2}\sum_{k=1}^{d'} (X_k')^2,$$

and, for $p' \in \mathcal{P}(\mathbb{R}^{N'})$, *determine* $t\rightsquigarrow\mathcal{L}''_{t,p'}$ *by*

$$\mathcal{L}''_{t,p'} = X_0''\big(p'(t)\big) + \frac{1}{2}\sum_{k=d'+1}^{d} X_k''\big(p'(t)\big)^2.$$

Then, the martingale problem for $\mathcal{L}'$ *starting at* x' *has exactly one solution* $\mathbb{P}_{x'}^{\mathcal{L}'}$, *and, for* $\mathbb{P}_x^{\mathcal{L}'}$*-almost every* $p' \in \mathcal{P}(\mathbb{R}^{N'})$, *the martingale problem for* $t\rightsquigarrow\mathcal{L}''_{t,p'}$ *has exactly one solution* $\mathbb{P}_{x''}^{\mathcal{L}'',p'}$ *starting at* x''. *In addition,* $p'\rightsquigarrow\mathbb{P}_{x''}^{\mathcal{L}'',p'}$ *is measurable on the set of* p'*'s at which* $\mathbb{P}_{x''}^{\mathcal{L}'',p'}$ *exists. Finally, if we write*

$$\mathcal{P}(\mathbb{R}^N) \ni p = (p',p'') \in \mathcal{P}(\mathbb{R}^{N'}) \times \mathcal{P}(\mathbb{R}^{N''}),$$

then

$$\mathbb{P}_x^{\mathcal{L}} = \int_{\mathcal{P}(\mathbb{R}^{N'})} \delta_{p'} \times \mathbb{P}_{x''}^{\mathcal{L}'';p'}\, \mathbb{P}_{x'}^{\mathcal{L}'}(dp'). \tag{3.30}$$

Equivalently,

$$\mathbb{P}_x^{\mathcal{L}}\big(A \times \mathcal{P}(\mathbb{R}^{N''})\big) = \mathbb{P}_{x'}^{\mathcal{L}'}(A), \quad A \in \mathcal{B}_{\mathcal{P}(\mathbb{R}^{N'})},$$

and, when $\mathcal{F}'$ *denotes the sub-*σ*-algebra of* $\mathcal{B}_{\mathcal{P}(\mathbb{R}^N)}$ *generated by* $\{p'(t) : t \geq 0\}$,

$$\mathbb{P}_x^{\mathcal{L}}(B \,|\, \mathcal{F}')(p') = \delta_{p'} \times \mathbb{P}_{x''}^{\mathcal{L}'';p'}(B) \quad \mathbb{P}_x^{\mathcal{L}}\text{-almost surely for each } B \in \mathcal{F}.$$

PROOF: The first step is to realize that, without loss in generality, we may assume that the X_k's have bounded derivatives of all orders. Indeed, by appealing to Corollary 3.11 in the same way as we did in §3.2.4, it is easy to reduce to the bounded case via a cut-off procedure which respects the decomposition of $\mathbb{R}^N$ into $\mathbb{R}^{N'} \times \mathbb{R}^{N''}$. Thus, we will assume that all derivatives of the X_k's are bounded.

The next step is to observe that, for each $\mathfrak{W}(\mathbb{R}^d) \ni \mathbf{w} = (\mathbf{w}', \mathbf{w}'') \in \mathfrak{W}(\mathbb{R}^{d'}) \times \mathfrak{W}(\mathbb{R}^{d''})$ (where $d'' = d - d'$),

$$p_n(t, x, \mathbf{w}) = \big(p_n'(t, x', \mathbf{w}'), p_n''(t, x'', (\mathbf{w}', \mathbf{w}''))\big),$$

where $p_n'(\,\cdot\,, x', \mathbf{w}') \in \mathcal{P}(\mathbb{R}^{N'})$ is determined from $(X_0', \dots, X_{d'}')$, x', and $\mathbf{w}'$ by the prescription in (2.17), and $p_n''(\,\cdot\,, x'', (\mathbf{w}', \mathbf{w}'')) \in \mathcal{P}(\mathbb{R}^{N''})$ is determined from

$$t \rightsquigarrow \Big(X_0''\big(p_n'(t, x', \mathbf{w}')\big), X_{d'+1}''\big(p_n'(t, x', \mathbf{w}')\big), \dots, X_d''\big(p_n'(t, x', \mathbf{w}')\big)\Big),$$

x'', and $\mathbf{w}''$ as in (3.23). This observation is nothing more than the remark which we made in the introduction about solving lower triangular systems of ordinary differential equations. Moreover, because $\mathbf{w}'$ is $\mu_{\mathbb{R}^d}$-independent of $\mathbf{w}''$, it is now clear that the $\mu_{\mathbb{R}^d}$-distribution $\mathbb{P}_n$ of $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto p_n(\,\cdot\,, x, \mathbf{w}) \in \mathcal{P}(\mathbb{R}^N)$ admits the decomposition

$$\mathbb{P}_n = \int_{\mathcal{P}(\mathbb{R}^{N'})} \delta_{p_n'(\,\cdot\,, x', \mathbf{w}')} \times \mathbb{P}_{n,\mathbf{w}'}''\, \mathbb{P}_n'(d\mathbf{w}'),$$

where $\mathbb{P}_n'$ is the $\mu_{\mathbb{R}^{d'}}$-distribution of $\mathbf{w}' \in \mathfrak{W}(\mathbb{R}^{d'}) \longmapsto p_n(\,\cdot\,, x', \mathbf{w}') \in \mathcal{P}(\mathbb{R}^{N'})$ and $\mathbb{P}_{n,\mathbf{w}'}''$ is the $\mu_{\mathbb{R}^{d''}}$-distribution of

$$\mathbf{w}'' \in \mathfrak{W}(\mathbb{R}^{d''}) \longmapsto p_n''(\,\cdot\,, x'', (\mathbf{w}', \mathbf{w}'')) \in \mathcal{P}(\mathbb{R}^{N''}).$$

Hence, the desired result follows from Theorems 2.40, 2.68, and 3.27 after one lets $n \to \infty$. □

3.5 Semigroups of Diffeomorphisms

Unlike classical flows generated by vector fields, stochastic flows, like the ones in § 2.4, seem to go in only one direction: forward in time. For this reason, they form *semigroups*, rather than groups, and so it is not immediately evident whether their members are diffeomorphic. The purpose of this section is to show that, although there are technical difficulties in doing so, the flows in § 2.4 can be reversed, and, as a consequence, we will know that they are semigroups of diffeomorphisms.

3.5.1. Flowing Backwards. Assume that the vector fields $(X_0, \dots, X_d)$ have bounded derivatives of all orders, and let $\mathbf{w} \rightsquigarrow p(\,\cdot\,, *\,, \mathbf{w})$ be determined accordingly, as in Theorems 2.40 and 2.58. At the same time, let $\mathbf{w} \rightsquigarrow \breve{p}(\,\cdot\,, *\,, \mathbf{w})$ be the map associated with the vector fields $(-X_0, \dots, -X_d)$. Finally, given $t \in (0, \infty)$, set

$$\breve{\mathbf{w}}^t(\tau) = \begin{cases} \mathbf{w}(t) - \mathbf{w}(t-\tau) & \text{when } \tau \in [0,t] \\ \mathbf{w}(\tau) & \text{when } \tau \in (t, \infty). \end{cases}$$

Using the characterization of Brownian motion in terms of independent, Gaussian increments, it is an easy matter to check that, for each t, the transformation $\mathbf{w} \rightsquigarrow \breve{\mathbf{w}}^t$ is $\mu_{\mathbb{R}^d}$-measure preserving. Hence, for each $t \in (0, \infty)$, $\mathbf{w} \rightsquigarrow \breve{p}(\,\cdot\,, *\,, \breve{\mathbf{w}}^t)$ is $\mu_{\mathbb{R}^d}$-well-defined, and, for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$,

$$\breve{p}_n(\,\cdot\,, *\,, \breve{\mathbf{w}}^t) \longrightarrow \breve{p}(\,\cdot\,, *\,, \breve{\mathbf{w}}^t) \quad \text{uniformly on compacts,}$$

where $\mathbf{w} \rightsquigarrow \breve{p}_n(\,\cdot\,, x, \breve{\mathbf{w}}^t)$ is given by the prescription in (2.17) with $(X_0, \dots, X_d)$ replaced by $(-X_0, \dots, -X_d)$ and $\mathbf{w}$ replaced by $\breve{\mathbf{w}}^t$.

The reason for introducing these considerations is that, for each $t \in [0, \infty)$,

$$(3.31)\quad p\big(t, \breve{p}(t, x, \breve{\mathbf{w}}^t), \mathbf{w}\big) = x = \breve{p}\big(t, p(t, x, \mathbf{w}), \breve{\mathbf{w}}^t\big) \quad \text{for } \mu_{\mathbb{R}^d}\text{-almost every } \mathbf{w}.$$

To prove (3.31), we will use the results of §2.4.5. Namely, given $t \in (0, \infty)$, define $\{\mathcal{T}_n : n \in \mathbb{N}\}$ so that $\mathcal{T}_n = \{\tau_{m,n} : m \in \mathbb{N}\}$, where $\tau_{0,n} = 0$, $\tau_{1,n} = t - [t]_n$,[7] and $\tau_{m+1,n} = 2^{-n} + \tau_{m,n}$ for $m \geq 1$. Then, by (2.59), we know that, for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$,

$$p_{\mathcal{T}_n}(\,\cdot\,, *\,, \mathbf{w}) \longrightarrow p(\,\cdot\,, *\,, \mathbf{w}) \quad \text{and} \quad \breve{p}_{\mathcal{T}_n}(\,\cdot\,, *\,, \breve{\mathbf{w}}^t) \longrightarrow \breve{p}(\,\cdot\,, *\,, \breve{\mathbf{w}}^t)$$

uniformly on compacts. Hence, (3.31) will follow if we can show that, for each $n \in \mathbb{N}$ and $\mathbf{w}$,

$$p_{\mathcal{T}_n}\big(\tau_{m,n}, \breve{p}_{\mathcal{T}_n}(\tau_{m,n}, x, \breve{\mathbf{w}}^{\tau_{m,n}}), \mathbf{w}\big) = x = \breve{p}_{\mathcal{T}_n}\big(\tau_{m,n}, p_{\mathcal{T}_n}(\tau_{m,n}, x, \mathbf{w}), \breve{\mathbf{w}}^{\tau_{m,n}}\big)$$
$$\text{for all } m \in \mathbb{N},$$

[7] We use $[r]$ to denote the integer part of $r \in \mathbb{R}$ and take $[r]_n \equiv 2^{-n}[2^n r]$.

and this can be done using the identities

$$
\begin{aligned}
E(\Xi,E(-\Xi,x)) = x &= E(-\Xi,E(\Xi,x)), && (\Xi,x)\in\mathbb{R}^{d+1}\times\mathbb{R}^N,\\
\big[\delta_s(\check{\mathbf{w}}^t)\big](\tau) &= \check{\mathbf{w}}^{t-s}(\tau), && \tau\in[0,t-s],\\
\Big[\big((\delta_s\mathbf{w})^\vee\big)^t\Big](\tau) &= \check{\mathbf{w}}^{t+s}(\tau), && \tau\in[0,t],
\end{aligned}
\tag{3.32}
$$

the flow property for $p_{\mathcal{T}_n}(\,\cdot\,,*,\mathbf{w})$ and $\check{p}_{\mathcal{T}_n}(\,\cdot\,,*,\mathbf{w})$ at times from $\mathcal{T}_n$, and induction on m.

3.5.2. Existence of a Continuous Version. [8] Although (3.31) is our basic result, we still want to improve it in two directions. In the first place, we want to show that it is possible to choose one $\mu_{\mathbb{R}^d}$-null set Λ such that

$$
\begin{gathered}
p\big(t,\check{p}(t,x,\check{\mathbf{w}}^t),\mathbf{w}\big) = x = \check{p}\big(t,p(t,x,\mathbf{w}),\check{\mathbf{w}}^t\big)\\
\text{for all } t\in[0,\infty) \text{ and } \mathbf{w}\notin\Lambda.
\end{gathered}
\tag{3.33}
$$

At the moment, (3.33) does not even make sense, because, for each $t\in[0,\infty)$, $\mathbf{w}\rightsquigarrow\check{p}(t,*,\check{\mathbf{w}}^t)$ is defined only up to a set of $\mu_{\mathbb{R}^d}$-measure 0. Nonetheless, starting from (3.31), it should be clear the (3.33) would both make sense and be true if we could choose a version of $\mathbf{w}\rightsquigarrow\check{p}(t,x,\check{\mathbf{w}}^t)$ for which

$$
(t,x)\in[0,\infty)\times\mathbb{R}^N \longmapsto \check{p}(t,x,\check{\mathbf{w}}^t)\in\mathbb{R}^N
$$

is $\mu_{\mathbb{R}^d}$-almost surely continuous.

Unfortunately, the proof that this is possible is not so easy and will require us to use *Kolmogorov's Continuity Criterion* which says (cf. Corollary 2.1.5 and Exercise 2.4.1 in [42]) that such a version exists if, for each $n\in\mathbb{N}$, there exists a $C_n<\infty$ such that

$$
\begin{gathered}
\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\big|\check{p}(t,y,\check{\mathbf{w}}^t)-\check{p}(s,x,\check{\mathbf{w}}^s)\big|_{\mathbb{R}^N}^{2N+2}\Big]^{\frac{1}{2N+2}} \le C_n\Big(|t-s|^{\frac12}+|y-x|_{\mathbb{R}^N}\Big)\\
\text{for all dyadic } 0\le s<t\le 2^n \text{ and all } x,\ y\in\mathbb{R}^N.
\end{gathered}
\tag{3.34}
$$

Actually, because $\mathbf{w}\rightsquigarrow\check{p}(\,\cdot\,,*,\check{\mathbf{w}}^t)$ has the same distribution as $\mathbf{w}\rightsquigarrow\check{p}(\,\cdot\,,*,\mathbf{w})$, it is easy to check that part of (3.34) holds. Namely, by (2.52)) applied (with $(-X_0,\dots,-X_d)$ in place of $(X_0,\dots,X_d)$), The Fundamental Theorem of Calculus, Jensen's Inequality, and Fubini's Theorem, there exists an $A_n<\infty$ such that, for any $(t,x,y)\in[0,2^n]\times\mathbb{R}^N\times\mathbb{R}^N$:

$$
\begin{aligned}
&\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\big|\check{p}(t,y,\check{\mathbf{w}}^t)-\check{p}(t,x,\check{\mathbf{w}}^t)\big|_{\mathbb{R}^N}^{2N+2}\Big]\\
&\quad\le |y-x|^{2N+2}\int_0^1 \mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\big\|J\check{p}(t,\theta x+(1-\theta)y,\mathbf{w})\big\|_{\mathrm{H.S.}}^{2N+2}\Big]\,d\theta \le A_n|y-x|_{\mathbb{R}^N}^{2N+2},
\end{aligned}
$$

[8] The material in this subsection is somewhat abstruse and will not be used in the sequel.

where $J\breve{p}(t, *, \mathbf{w}) = \frac{\partial \breve{p}}{\partial x}(t, *, \mathbf{w})$ is the Jacobian matrix of $\breve{p}(t, *, \mathbf{w})$. Hence, the verification of (3.34) reduces to checking that there is a $B_n < \infty$ such that

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\big|\breve{p}(t, x, \breve{\mathbf{w}}^t) - \breve{p}(s, x, \breve{\mathbf{w}}^s)\big|_{\mathbb{R}^N}^{2N+2}\Big] \le B_n(t-s)^{N+1} \quad \text{for all dyadic } 0 \le s < t \le 2^n. \tag{3.35}$$

The key to our proof of (3.35) is the flow property

$$\breve{p}\big(t, x, \breve{\mathbf{w}}^t\big) = \breve{p}\big(s, \breve{p}(t-s, x, \breve{\mathbf{w}}^t), \breve{\mathbf{w}}^s\big), \tag{3.36}$$

which holds $\mu_{\mathbb{R}^d}$-almost surely for each dyadic pair (s, t) with $0 \le s < t < \infty$. The verification of (3.36) is straight-forward. Indeed, choose $n_0 \in \mathbb{N}$ so that both $2^{n_0}s$ and $2^{n_0}t$ are integers, and, for $n \ge n_0$, use the definition of $\breve{p}_n$ to see that (3.36) holds when $\breve{p}$ is replaced by $\breve{p}_n$ throughout. To use (3.36), observe that $\mathbf{w} \rightsquigarrow \breve{p}(t-s, x, \breve{\mathbf{w}}^t)$ is $\sigma(\{\mathbf{w}(t) - \mathbf{w}(\tau) : \tau \in [s, t]\})$-measurable whereas $\mathbf{w} \rightsquigarrow \breve{p}(s, x, \breve{\mathbf{w}}^s)$ is $\sigma(\{\mathbf{w}(\tau) : \tau \in [0, s]\})$-measurable. Hence, referring to the preceding paragraph, we find that

$$\begin{aligned}
&\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\big|\breve{p}(t, x, \breve{\mathbf{w}}^t) - \breve{p}(s, x, \breve{\mathbf{w}}^s)\big|_{\mathbb{R}^N}^{2N+2}\Big] \\
&\quad = \mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\big|\breve{p}\big(s, \breve{p}(t-s, x, \breve{\mathbf{w}}^t), \breve{\mathbf{w}}^s\big) - \breve{p}(s, x, \breve{\mathbf{w}}^s)\big|_{\mathbb{R}^N}^{2N+2}\Big] \\
&\quad \le A_n \mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\big|\breve{p}(t-s, x, \mathbf{w}) - x\big|_{\mathbb{R}^N}^{2N+2}\Big],
\end{aligned}$$

which, when combined with (2.44), yields (3.35).

Given (3.34), Kolmogorov's Continuity Criterion tells us that it is possible to construct a measurable

$$\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto q(\,\cdot\,, *, \mathbf{w}) \in C\big([0,\infty) \times \mathbb{R}^N; \mathbb{R}^N\big)$$

so that $q(t, *, \mathbf{w}) = \breve{p}(t, *, \breve{\mathbf{w}}^t)$ (a.s, $\mu_{\mathbb{R}^d}$) for every dyadic $t \in [0, \infty)$. Hence, (3.31) plus continuity imply (3.33) in the sense that there is a $\mu_{\mathbb{R}^d}$-null set Λ such that

$$p\big(t, q(t, x, \mathbf{w}), \mathbf{w}\big) = x = q\big(t, p(t, x, \mathbf{w}), \mathbf{w}\big) \quad \text{for all } t \in [0, \infty) \text{ and } \mathbf{w} \notin \Lambda.$$

In particular, this shows that, for

$$\mathbf{w} \notin \Lambda \implies t \in [0, \infty) \longmapsto p(t, *, \mathbf{w}) \in \text{Homeo}(\mathbb{R}^N) \text{ continuously}, \tag{3.37}$$

where $\text{Homeo}(\mathbb{R}^N)$ is the group of homeomorphisms from $\mathbb{R}^N$ onto itself with the topology induced by convergence in $C(\mathbb{R}^N; \mathbb{R}^N)$ of both the map and its inverse.

3.5.3. Non-Degenerate Jacobian. Let $\mathrm{Diff}(\mathbb{R}^N) \subsetneq C^\infty(\mathbb{R}^N;\mathbb{R}^N)$ denote the group of diffeomorphisms from $\mathbb{R}^N$ onto itself, and endow $\mathrm{Diff}(\mathbb{R}^N)$ with the topology which is induced by convergence in $C^\infty(\mathbb{R}^N;\mathbb{R}^N)$ of both the function and its inverse. As an application of the Implicit Function Theorem, one knows that if $t \in [0,\infty) \longmapsto F_t \in C^\infty(\mathbb{R}^N;\mathbb{R}^N)$ is continuous and, for each $t \in [0,\infty)$, $F_t \in \mathrm{Homeo}(\mathbb{R}^N)$ and the Jacobian matrix $JF_t(x) \equiv \frac{\partial F_t}{\partial x}(x)$ is non-singular for all $x \in \mathbb{R}^N$, then $F_t \in \mathrm{Diff}(\mathbb{R}^N)$ for all $t \in [0,\infty)$ and $t \in [0,\infty) \longmapsto F_t \in \mathrm{Diff}(\mathbb{R}^N)$ is continuous. Thus, by combining Theorem 2.58 with (3.37), we will know that there is a $\mu_{\mathbb{R}^d}$-null set Λ such that

$$\text{(3.38)} \qquad \mathbf{w} \notin \Lambda \implies t \in [0,\infty) \longmapsto p(t,\,*\,,\mathbf{w}) \in \mathrm{Diff}(\mathbb{R}^N) \text{ continuously}$$

once we show that, for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$, the Jacobian matrix $Jp(t,x,\mathbf{w}) = \frac{\partial p(t,\,*\,,\mathbf{w})}{\partial x}(x)$ is non-singular for all $(t,x) \in [0,\infty) \times \mathbb{R}^N$.

To test the non-singularity of $Jp(t,x,\mathbf{w})$, set $D_n(t,x,\mathbf{w})$ equal to the determinant of the Jacobian matrix (cf. (2.17)) $Jp_n(t,x,\mathbf{w})$. Because, by the elementary theory of ordinary differential equations,

$$\det\left(\frac{\partial E(\Xi,\,*\,)}{\partial x}(x)\right) = \exp\left(\int_0^1 (\mathrm{div} X_\Xi)\big(E(\tau\Xi,x)\big)\,d\tau\right),$$

where $\mathrm{div}Y = \sum_{i=1}^N \partial_i Y^i$ denotes the standard Euclidean divergence of the vector field Y, we see that $D_n(t,x,\mathbf{w}) > 0$ and that

$$\begin{aligned} L_n(t,x,\mathbf{w}) &\equiv \log D_n(t,x,\mathbf{w}) \\ &= \int_0^1 \left(\sum_{m=0}^\infty (\mathrm{div}X_{\Xi_{(m,n)}(t,\mathbf{w})})\Big(E\big(\tau\Xi_{m,n}(t,\mathbf{w}), p_n(T_{m,n},x,\mathbf{w})\big)\Big)\right)\,d\tau. \end{aligned}$$

Since $D_n(\,\cdot\,,\,*\,,\mathbf{w}) \longrightarrow D(\,\cdot\,,\,*\,,\mathbf{w}) \equiv \det\big(Jp(t,x,\mathbf{w})\big)$ in $C([0,\infty)\times\mathbb{R}^N;\mathbb{R})$ for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$, it suffices for us to show that that $\lim_{n\to\infty} L_n(\,\cdot\,,\,*\,,\mathbf{w})$ exists in $C([0,\infty)\times\mathbb{R}^N;\mathbb{R})$ for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$. Starting from the preceding expression for $L_n(t,x,\mathbf{w})$, one can prove that this limit exists by a repetition of the same sort of argument as we used in §2.3.2. However, there is an alternative, and preferable, route. Namely, define

$$\text{(3.39)} \quad (\tilde{X}_k)_{(x^0,x)} = \big[\mathrm{div}(X_k)\big](x)\partial_0 + \sum_{i=1}^N X_k^i(x)\partial_{x^i} \quad \text{and } \tilde{\mathcal{L}} = \tilde{X}_{x^0} + \frac{1}{2}\sum_{k=1}^d \tilde{X}_k^2$$

on $\mathbb{R}^{N+1} = \mathbb{R}\times\mathbb{R}^N$; and, for a given $\tilde{x} \in \mathbb{R}^{N+1}$, consider the sequence of paths $\{\tilde{p}_n(\,\cdot\,,\tilde{x},\mathbf{w}) : n \in \mathbb{N}\}$ associated with $(\tilde{X}_0,\dots,\tilde{X}_d)$, as in (2.17). It is then an easy matter to identify $L_n(t,x,\mathbf{w})$ with the 0th component $\tilde{p}_n^0\big(t,(0,x),\mathbf{w}\big)$ of $\tilde{p}_n\big(t,(0,x),\mathbf{w}\big)$. Hence, we have now proved the following interesting fact.

3.40 THEOREM. *Referring to the preceding discussion, let*

$$\mathbf{w}\rightsquigarrow\tilde{p}(\,\cdot\,,*\,,\mathbf{w}) = \big(\tilde{p}(\,\cdot\,,*\,,\mathbf{w})^0,\dots,\tilde{p}(\,\cdot\,,*\,,\mathbf{w})^N\big) \in C^{0,\infty}([0,\infty)\times\mathbb{R}^{N+1};\mathbb{R}^{N+1})$$

be the map associated with $(\tilde{X}_0,\dots,\tilde{X}_d)$ *as in Theorems 2.40 and 2.58. Then, for* $\mu_{\mathbb{R}^d}$*-almost every* $\mathbf{w}$,

$$\det\left(\frac{\partial p(t,*\,,\mathbf{w})}{\partial x}(x)\right) = \exp\Big(\tilde{p}(t,(0,x),\mathbf{w})^0\Big), \quad (t,x)\in[0,\infty)\times\mathbb{R}^N. \tag{3.41}$$

In particular, there is a $\mu_{\mathbb{R}^d}$*-null set* Λ *for which (3.38) holds.*

3.5.4. In General. The second improvement alluded to at the beginning of §3.5.2 entails the replacement of the boundedness condition on our vector fields by the assumption that the martingale problems for $\mathcal{L}$ is well-posed. As our first attempt in this direction, we have the following general statement.

3.42 THEOREM. *Assume that the martingale problem for* $\mathcal{L}$ *is well-posed. Then, for* $\mu_{\mathbb{R}^d}$*-almost every* $\mathbf{w}$ *and every* $t\in[0,\infty)$, $p(t,*\,,\mathbf{w})$ *is a one-to-one,* C^∞ *map on* $\mathbb{R}^N$. *Moreover, the martingale problem for (cf. (3.39))* $\tilde{\mathcal{L}}$ *is well-posed and (3.41) holds (a.s.,* $\mu_{\mathbb{R}^d}$*). In particular, the Jacobian of* $p(t,*\,,\mathbf{w})$ *is non-singular for all* $t\geq 0$ *and* $\mu_{\mathbb{R}^d}$*-almost all* $\mathbf{w}$. *Thus, for each* $t\in(0,\infty)$,

$$p(t,\mathbb{R}^N,\mathbf{w}) = \mathbb{R}^N \text{ (a.s., } \mu_{\mathbb{R}^d}) \iff p(t,*\,,\mathbf{w})\in \mathrm{Diff}(\mathbb{R}^N) \quad \text{(a.s., } \mu_{\mathbb{R}^d});$$

and

$$\bigcap_{t\geq 0} p(t,\mathbb{R}^N,\mathbf{w}) = \mathbb{R}^N \text{ (a.s., } \mu_{\mathbb{R}^d}) \iff$$

$$t\in[0,\infty)\longmapsto p(t,*\,,\mathbf{w})\in\mathrm{Diff}(\mathbb{R}^N) \text{ is continuous (a.s., } \mu_{\mathbb{R}^d}).$$

PROOF: By Theorem 3.15, we know that (3.12) holds for every $x\in\mathbb{R}^N$. Thus, since (cf. §3.2.1) each $\zeta^m(*\,,\mathbf{w})$ is lower semi-continuous and $\zeta^m(*\,,\mathbf{w})\nearrow \mathfrak{e}(*\,,\mathbf{w})$ for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$, we know that, for each $T\in(0,\infty)$,

$$\begin{gathered} V_m(T,\mathbf{w}) \equiv \{x:\ \zeta^m(x,\mathbf{w})>T\} \\ \text{is open for every } \mathbf{w} \text{ and } V_m(T,\mathbf{w})\nearrow\mathbb{R}^N \quad \text{for } \mu_{\mathbb{R}^d}\text{-almost every } \mathbf{w}. \end{gathered} \tag{3.43}$$

In particular, since

$$p(\,\cdot\,,*\,,\mathbf{w})\restriction[0,T]\times V_m(T,\mathbf{w}) = p^m(\,\cdot\,,*\,,\mathbf{w})\restriction[0,T]\times V_m(T,\mathbf{w}) \quad \text{(a.s., } \mu_{\mathbb{R}^d}),$$

the first assertion follows easily from the fact that, for each m, $t\in[0,\infty)\longmapsto p^m(t,*\,,\mathbf{w})\in\mathrm{Diff}(\mathbb{R}^N)$ is continuous. Moreover, given this observation, it follows from (3.41), applied to $\tilde{p}^m(\,\cdot\,,*\,,\mathbf{w})$, that $\tilde{p}(\,\cdot\,,(0,y),\mathbf{w})^0$ almost never explodes for any $y\in\mathbb{R}^N$. Hence, since

$$\tilde{p}(\,\cdot\,,(y^0,y),\mathbf{w}) = \Big(y^0+\tilde{p}(\,\cdot\,,(0,y),\mathbf{w})^0, p(\,\cdot\,,y,\mathbf{w})\Big) \quad \mu_{\mathbb{R}^d}\text{-almost surely,}$$

it follows that the martingale problem for $\tilde{\mathcal{L}}$ is also well-posed. At the same time, (3.41) for the present case follows immediately from (3.41) applied to p^m and $(\tilde{p}^m)^0$.

Given the first assertion, the rest of the theorem is a more or less trivial application of the Implicit Function Theorem. □

3.44 COROLLARY. *Suppose that the martingale problems both for $\mathcal{L}$ and for*

$$\check{\mathcal{L}} \equiv -X_0 + \frac{1}{2}\sum_{k=1}^{d} X_k^2 \tag{3.45}$$

are well-posed. Then, for each $t \in [0,\infty)$, $p(t,*,\mathbf{w}) \in \mathrm{Diff}(\mathbb{R}^N)$ and (3.31) holds $\mu_{\mathbb{R}^d}$-almost surely.

PROOF: In view of Theorem 3.42 and (3.31) for p^m and $\check{p}^m$ replacing p and $\check{p}$, all that we have to do is check that $p(t,\mathbb{R}^N,\mathbf{w}) = \mathbb{R}^N$ $\mu_{\mathbb{R}^d}$-almost surely. To this end, define $\check{V}_m(t,\mathbf{w})$ as in (3.43), only this time relative to $(-X_0,\dots,-X_d)$ instead of $(X_0,\dots,X_d)$. Then, $\check{V}_m(t,\check{\mathbf{w}}^t) \nearrow \mathbb{R}^N$ $\mu_{\mathbb{R}^d}$-almost surely. Moreover, for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$,

$$p\big(t,\check{p}^m(t,y,\check{\mathbf{w}}^t),\mathbf{w}\big) = y = p^m\big(t,\check{p}^m(t,y,\check{\mathbf{w}}^t),\mathbf{w}\big), \quad y \in \check{V}_m(t,\check{\mathbf{w}}^t).$$

Thus, we are done. □

3.6 Invariant and Symmetric Measures

Let $(X_0,\dots,X_d)$ be smooth vector fields, form the operator $\mathcal{L}$ as in (2.3), and assume that the martingale problem for $\mathcal{L}$ is well-posed. Next, define the operators $\{\mathbf{P}_t^{\mathcal{L}} : t > 0\}$ on $C_\mathrm{b}(\mathbb{R}^N;\mathbb{R})$ by (cf. Theorem 3.7)

$$\big[\mathbf{P}_t^{\mathcal{L}} f\big](x) = \mathbb{E}^{\mathbb{P}_x^{\mathcal{L}}}\Big[f\big(p(t)\big)\Big] = \mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[f\big(p(t,x,\mathbf{w})\big)\Big]. \tag{3.46}$$

Then we know that each $\mathbf{P}_t^{\mathcal{L}}$ is a non-negativity preserving map of $C_\mathrm{b}(\mathbb{R}^N;\mathbb{R}^N)$ into itself and that $\mathbf{P}_t^{\mathcal{L}}$ leaves the constant function $\mathbf{1}$ invariant. In addition, again by Theorem 3.7, we know that $\{\mathbf{P}_t^{\mathcal{L}} : t > 0\}$ forms a semigroup. In the vernacular of probability theory, these conclusions are summarized by saying that $\{\mathbf{P}_t^{\mathcal{L}} : t > 0\}$ is a *Feller continuous* (because it preserves $C_\mathrm{b}(\mathbb{R}^N;\mathbb{R}^N)$), *Markov* (because it preserves both non-negativity and constancy of functions) semigroup.

Notice that a Markov semigroup has a *dual* action on the space $\mathbf{M}_\mathrm{f}(\mathbb{R}^N)$ of totally finite, Borel, signed measures. Namely,

$$\nu \in \mathbf{M}_\mathrm{f}(\mathbb{R}^N) \longmapsto \nu\mathbf{P}_t^{\mathcal{L}} \quad \text{where } (\nu\mathbf{P}_t^{\mathcal{L}}, f) = (\nu, \mathbf{P}_t^{\mathcal{L}} f) \tag{3.47}$$

and we have used the functional analytic notation

$$(\nu, f) \equiv \int f\, d\nu$$

in order to highlight the role of duality here. Obviously, the Markov property is manifested in the dual operation by the fact that it preserves the space $\mathbf{M}_1(\mathbb{R}^N)$ of Borel probability measures on $\mathbb{R}^N$. In addition, it is clear that we can take advantage of the Markov property to define $\nu\mathbf{P}_t^{\mathcal{L}}$ for $\nu \in \mathbf{M}^+(\mathbb{R}^N)$, the cone of non-negative, Borel measures on $\mathbb{R}^N$. We simply take $\nu \in \mathbf{M}^+(\mathbb{R}^N) \longmapsto \nu\mathbf{P}_t^{\mathcal{L}} \in \mathbf{M}^+(\mathbb{R}^N)$ so that

$$\int f\, d(\nu\mathbf{P}_t^{\mathcal{L}}) = \int \mathbf{P}_t^{\mathcal{L}} f\, d\nu \quad \text{for } f \in C(\mathbb{R}^N;[0,\infty)).$$

For anyone brought up in the functional analytic tradition, the preceding comments suggest that one should seek elements ν of $\mathbf{M}_1(\mathbb{R}^N)$, or, at worst, $\mathbf{M}^+(\mathbb{R}^N)$, which are invariant measure under the dual action of $\{\mathbf{P}_t^{\mathcal{L}} : t > 0\}$. Indeed, we know that $\mathbf{1}$ is a simultaneous eigenvector with eigenvalue 1 for $\{\mathbf{P}_t^{\mathcal{L}} : t > 0\}$, and so we should investigate whether the dual action also admits a simultaneous eigenvector with eigenvalue 1. Of course, the present situation is well outside the scope of general (e.g., Frobenius-type) functional analytic theory, but one still can hope that something can be said. In particular, it is important to see to what extent the following, formal line of reasoning can be justified.

Suppose that ν is a locally finite, $\{\mathbf{P}_t^{\mathcal{L}} : t > 0\}$-invariant element of $\mathbf{M}^+(\mathbb{R}^N)$. Then one should be able to justify

$$\int \mathcal{L}f\, d\nu = \frac{d}{dt}\int \mathbf{P}_t^{\mathcal{L}} f\, d\nu\bigg|_{t=0} = 0$$

for a reasonable class of test functions f. Hence, thinking in terms of L. Schwartz's theory of distributions, one finds that $\mathcal{L}^*\nu = 0$, where

$$\mathcal{L}^* \equiv X_0^* + \frac{1}{2}\sum_{k=1}^{d}(X_k^*)^2 \quad \text{with } X_k^* \equiv -X_k - \operatorname{div}(X_k) \tag{3.48}$$

is the formal adjoint of $\mathcal{L}$. Conversely, and more significantly, one should hope that this line of reasoning will yield a criterion for recognizing when a ν is $\{\mathbf{P}_t^{\mathcal{L}} : t > 0\}$-invariant. That is, one hopes that

$$\mathcal{L}^*\nu = 0 \implies \nu = \nu\mathbf{P}_t^{\mathcal{L}} \quad \text{for all } t > 0,$$

at least when ν admits a smooth density (i.e., Radon–Nikodym derivative) with respect to Lebesgue measure.

3.6.1. Criterion for Invariance. Let $\mathcal{L}$ be defined, as in (2.3), from smooth vector fields $(X_0, \dots, X_d)$ on $\mathbb{R}^N$, determine the vector fields $(\hat{X}_0, \dots, \hat{X}_d)$ on $\mathbb{R}^{N+1} = \mathbb{R} \times \mathbb{R}^N$ by

$$\big(\hat{X}_k\big)_{\hat{x}} = -\big[\operatorname{div}(X_k)\big](x)\partial_{x^0} - X_k = -\big[\operatorname{div}(X_k)\big](x)\partial_{x^0} - \sum_{i=1}^{N} X_k^i(x)\partial_i, \tag{3.49}$$

and define $\hat{\mathcal{L}}$ on $C^2(\mathbb{R}^{N+1};\mathbb{R})$, accordingly. Everything in this section rests on the following corollary of the results in §3.5.4.

3.50 THEOREM. *Assume that the martingale problems for both $\mathcal{L}$ and (cf. (3.45)) $\check{\mathcal{L}}$ are well-posed. Then the martingale problem for $\hat{\mathcal{L}}$ is also well-posed. Moreover, if f and g are non-negative, measurable functions on $\mathbb{R}^N$, then*

$$\int_{\mathbb{R}^N} g(x)\,[\mathbf{P}_t^{\mathcal{L}} f](x)\,dx = \int_{\mathbb{R}^N} f(y)\,[\mathbf{P}_t^{\hat{\mathcal{L}}}\hat{g}]((0,y))\,dy, \tag{3.51}$$

where $\mathbf{P}_t^{\hat{\mathcal{L}}}$ is defined in terms of $\hat{\mathcal{L}}$ as in (3.46) and $\hat{g}((y^0,y)) \equiv e^{y^0} g(y)$.

PROOF: The first assertion follows immediately from Theorem 3.42 applied with the roles of $\mathcal{L}$ and $\check{\mathcal{L}}$ reversed. Next, let $\check{D}(t,*,\mathbf{w})$ denote the determinant of the Jacabian of $\check{p}(t,*,\mathbf{w})$. Then, because $\check{p}(t,*,\check{\mathbf{w}}^t)$ is, $\mu_{\mathbb{R}^d}$-almost surely, the inverse of $p(t,*,\mathbf{w})$, Jacobi's change of variable formula tells us that

$$\int_{\mathbb{R}^N} g(x)\,f(p(t,x,\mathbf{w}))\,dx = \int_{\mathbb{R}^N} f(y)\,g(\check{p}(t,y,\check{\mathbf{w}}^t))\check{D}(t,x,\check{\mathbf{w}}^t)\,dy \quad (\text{a.s.}, \mu_{\mathbb{R}^d}).$$

Now integrate both sides with respect to $\mu_{\mathbb{R}^d}$ and use (3.41), with $\hat{p}$ in place of $\check{p}$, to arrive at (3.51). □

To see how (3.51) gets applied, observe that (cf. (3.48))

$$[\hat{\mathcal{L}}\hat{g}]((y^0,y)) = e^{y^0}[\mathcal{L}^* g](y). \tag{3.52}$$

Next, suppose that $\rho \in C^\infty(\mathbb{R}^N;[0,\infty))$ satisfies $\mathcal{L}^*\rho \le 0$. Then, $\hat{\mathcal{L}}\hat{\rho} \le 0$, and so (cf. §3.2.1 and apply the results there to the "hatted" quantities here)

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\hat{\rho}\big(\hat{p}(t\wedge\hat{\zeta}^m(\hat{y},\mathbf{w}),\hat{y},\mathbf{w})\big)\Big] \le \hat{\rho}(\hat{y}),$$

which, by Fatou's Lemma, gives

$$[\mathbf{P}_t^{\hat{\mathcal{L}}}\hat{\rho}](\hat{y}) = \mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\hat{\rho}\big(\hat{p}(t,\hat{y},\mathbf{w})\big)\Big] \le \hat{\rho}(\hat{y}), \quad (t,\hat{y}) \in [0,\infty)\times\mathbb{R}^{N+1},$$

when $m \to \infty$.

3.53 COROLLARY. *Let $\rho \in C^\infty(\mathbb{R}^N;[0,\infty))$ be given, and define $\nu_\rho(dy) = \rho(y)\,dy$. Under the conditions in Theorem 3.50, $\mathcal{L}^*\rho \le 0$ implies that*

$$\nu_\rho \mathbf{P}_t^{\mathcal{L}} \le \nu_\rho \quad \text{for all } t \in (0,\infty). \tag{3.54}$$

Moreover, if $\nu_\rho \in \mathbf{M}_1(\mathbb{R}^N)$, then[9]

$$\mathcal{L}^*\rho \le 0 \iff \nu_\rho = \nu_\rho \mathbf{P}_t^{\mathcal{L}} \text{ for all } t \in (0,\infty) \iff \mathcal{L}^*\rho = 0.$$

[9] This result is considerably less than optimal. In fact, as an application of the main result in [10], one knows that it suffices to assume that the martingale problem for $\mathcal{L}$ is well-posed.

PROOF: In view of the preceding paragraph and Theorem 3.50, the first part is already proved. Thus, suppose that $\nu_\rho \in \mathbf{M}_1(\mathbb{R}^N)$ and that $\mathcal{L}^*\rho \leq 0$. Then, for any $t \in (0, \infty)$, $\nu_\rho \geq \nu_\rho \mathbf{P}_t^{\mathcal{L}}$. At the same time, both ν_ρ and $\nu_\rho \mathbf{P}_t^{\mathcal{L}}$ are probability measures. Thus, the inequality must be an equality. Next, assume that $\nu_\rho \in \mathbf{M}_1(\mathbb{R}^N)$ and that $\nu_\rho = \nu_\rho \mathbf{P}_t^{\mathcal{L}}$ for all $t > 0$. Then, for any $f \in C_c^\infty(\mathbb{R}^N; \mathbb{R})$, Lebesgue's Dominated Convergence Theorem justifies

$$\int_{\mathbb{R}^N} \rho(x)\mathcal{L}f(x)\,dx = \lim_{t \searrow 0} \frac{1}{t} \int_{\mathbb{R}^N} \left(\mathbf{P}_t^{\mathcal{L}} f - f\right) d\nu_\rho = 0.$$

In other words, $\mathcal{L}^*\rho = 0$ in the sense of Schwartz distributions. But ρ is smooth, and therefore the $\mathcal{L}^*\rho = 0$ in the classical sense. $\square$

Although we will not enter into a lengthy discussion of the important role which invariant measures play in the analysis of Markov processes, the reader should be aware that, given a $\{\mathbf{P}_t^{\mathcal{L}} : t > 0\}$-invariant $\nu \in \mathbf{M}^+(\mathbb{R}^N)$, there are several general properties which can be immediately exploited. For one thing, an application of Jensen's inequality shows that, for each $q \in [1, \infty)$, $t \in (0, \infty)$, and $f \in C(\mathbb{R}^N; [0, \infty))$:

$$\left\|\mathbf{P}_t^{\mathcal{L}} f\right\|_{L^q(\nu;\mathbb{R})} \leq \left\|\mathbf{P}_t^{\mathcal{L}} f^q\right\|_{L^1(\nu;\mathbb{R})}^{\frac{1}{q}} = \|f\|_{L^q(\nu;\mathbb{R})}. \tag{3.55}$$

Thus, at least if ν is locally finite, it is easy to see that $\{\mathbf{P}_t^{\mathcal{L}} : t > 0\}$ determines a unique, strongly continuous semigroup of contractions on $L^q(\nu; \mathbb{R})$.

To go in a more measure theoretic direction, suppose that $\nu \in \mathbf{M}_1(\mathbb{R}^N)$ is invariant. Then standard methods (e.g., Kolmogorov's Extension Theorem) allow one to construct a *stationary measure* $\mathbb{P}$ on $C(\mathbb{R}; \mathbb{R}^N)$ with the properties that, for any $s \in \mathbb{R}$, the $\mathbb{P}$-distribution of $p \in C(\mathbb{R}; \mathbb{R}^N) \longmapsto p(s) \in \mathbb{R}^N$ is ν and the conditional distribution of

$$p \in C(\mathbb{R}; \mathbb{R}^N) \longmapsto p(s + \cdot\,) \restriction [0, \infty) \in \mathcal{P}(\mathbb{R}^N) \quad \text{given } \sigma\big(\{p(\tau) : \tau \in (-\infty, s]\}\big)$$

is $\mathbb{P}_{p(s)}^{\mathcal{L}}$. In particular, $\mathbb{P}$ is time-shift invariant, and one can start studying its ergodic properties.

3.6.2. Symmetric Measures. There are times when an invariant measure $\nu \in \mathbf{M}^+(\mathbb{R}^N)$ is better than invariant. To wit, it may be *symmetric* in the sense that

$$\int_{\mathbb{R}^N} g\,\mathbf{P}_t^{\mathcal{L}} f\,d\nu = \int_{\mathbb{R}^N} f\,\mathbf{P}_t^{\mathcal{L}} g\,d\nu \quad \text{for all } t \in (0, \infty) \text{ and } f \in C\big(\mathbb{R}^N; [0, \infty)\big).$$

Trivially, by taking $g = \mathbf{1}$, one sees that such a ν is invariant. However, except in rare circumstances, not every invariant measure is symmetric. In fact, a little experience in such matters is enough to convince one that the existence of a symmetric measure is a rather rare event. On the other hand, one can force the issue, and, in the present setting, one way to force it is the following. Start with smooth vector fields $(X_1, \dots, X_d)$ on $\mathbb{R}^N$. Next, given a smooth $\rho : \mathbb{R}^N \longrightarrow (0,\infty)$, define (cf. (3.48))

$$(3.56) \qquad \mathcal{L}_\rho f = -\frac{1}{2\rho}\sum_{k=1}^{d} X_k^*(\rho X_k f), \quad f \in C^2(\mathbb{R}^N;\mathbb{R}).$$

Clearly, by taking

$$X_0 = \frac{1}{2\rho}\sum_{k=1}^{d} \operatorname{div}(\rho X_k) X_k,$$

one can re-write $\mathcal{L}_\rho$ in the form given on the right side of (2.3). In particular, the theory which we have been developing applies to $\mathcal{L}_\rho$. On the other hand, it is best to leave $\mathcal{L}_\rho$ in its original form when verifying the identity

$$(3.57) \qquad \mathcal{L}_\rho^*(\rho g) = \rho \mathcal{L}_\rho g, \quad g \in C^2(\mathbb{R}^N;\mathbb{R}).$$

3.58 THEOREM. *If the martingale problem for $\mathcal{L}_\rho$ is well-posed, then (cf. Corollary 3.53)*

$$(3.59) \qquad \int_{\mathbb{R}^N} g\, \mathbf{P}_t^{\mathcal{L}_\rho} f \, d\nu_\rho = \int_{\mathbb{R}^N} f\, \mathbf{P}_t^{\mathcal{L}_\rho} g \, d\nu_\rho, \quad t \in (0,\infty),$$

for all non-negative f, $g \in C(\mathbb{R}^N;\mathbb{R})$.

PROOF: Assume, for the moment, that the X_k's vanish off of some compact set. Proceeding via Theorem 3.50, we see that, for any pair f, $g \in C_c^\infty(\mathbb{R}^N;[0,\infty))$ and $t > 0$:

$$\int_{\mathbb{R}^N} g\, \mathbf{P}_t^{\mathcal{L}_\rho} f \, d\nu_\rho = \int_{\mathbb{R}^N} f(y) \big[\mathbf{P}_t^{\widehat{\mathcal{L}_\rho}} \widehat{\rho g}\big]\big((0,y)\big)\, dy.$$

Hence, (3.59) will follow when we check that

$$\big[\mathbf{P}_t^{\widehat{\mathcal{L}_\rho}} \widehat{\rho g}\big](\hat{y}) = e^{y^0}\rho(y)\big[\mathbf{P}_t^{\mathcal{L}_\rho} g\big](y) \quad \text{for } \hat{y} = (y^0, y) \in \mathbb{R}\times\mathbb{R}^N.$$

To this end, let $u(t,\hat{y})$ denote the right hand side of the preceding. Then, by (2.67) (applied with $\mathcal{L} = \widehat{\mathcal{L}_\rho}$), (3.52), and (3.57), one sees that $\frac{\partial u}{\partial t} = \widehat{\mathcal{L}_\rho} u$ on $[0,\infty)\times\mathbb{R}^{N+1}$; which, in conjunction with the last part of Theorem 2.40, means that

$$\tau \rightsquigarrow u\Big(t - t\wedge\tau, \hat{p}(t\wedge\tau, \hat{y}, \mathbf{w})\Big)$$

is a $\mu_{\mathbb{R}^d}$-martingale. In particular, the desired equality comes from equating the $\mu_{\mathbb{R}^d}$-expectation values of this martingale at $\tau = 0$ and $\tau = t$.

Finally, to remove the assumption that the X_k's are compactly supported, choose $\psi \in C_c^\infty(\mathbb{R}^N;[0,1])$ so that $\psi \equiv 1$ on $B_{\mathbb{R}^N}(\mathbf{0},1)$, set $(X_{k,m})_x = \psi\left(\frac{x}{m}\right)(X_k)_x$, define $\mathcal{L}_{\rho,m}$ and $\mathbf{w} \rightsquigarrow p^m(\,\cdot\,,*,\mathbf{w})$ from ρ and $(X_{0,m},\dots,X_{d,m})$

accordingly, and use $\zeta^m(x,\mathbf{w})$ to denote the first time $p^m(\,\cdot\,,x,\mathbf{w})$ leaves the ball $B_{\mathbb{R}^N}(\mathbf{0},m)$. Then,

$$\zeta^m(x,\mathbf{w}) \nearrow \infty \text{ and } p(\,\cdot\,,x,\mathbf{w}) \restriction [0,\zeta^m(x,\mathbf{w})) = p^m(\,\cdot\,,x,\mathbf{w}) \restriction [0,\zeta^m(x,\mathbf{w}))$$

$\mu_{\mathbb{R}^d}$-almost surely. Hence, for any $f,\,g \in C_c(\mathbb{R}^N;[0,\infty))$,

$$\begin{aligned}\int f\,\mathbf{P}_t^{\mathcal{L}_\rho} g\,d\nu_\rho &= \lim_{m\to\infty}\int f\,\mathbf{P}_t^{\mathcal{L}_{\rho,m}} g\,d\nu_\rho \\ &= \lim_{m\to\infty}\int g\,\mathbf{P}_t^{\mathcal{L}_{\rho m}} f\,d\nu_\rho = \int g\,\mathbf{P}_t^{\mathcal{L}_\rho} f\,d\nu_\rho. \quad \square\end{aligned}$$

As we said before, symmetric measures are even better than invariant ones. To give a hint about why, let $\{\bar{\mathbf{P}}_t^{\mathcal{L}_\rho} : t > 0\}$ denote the semigroup on $L^2(\nu_\rho;\mathbb{R})$, alluded to in the discussion following Corollary 3.53. It is then clear that each $\bar{\mathbf{P}}_t^{\mathcal{L}}$ is a self-adjoint contraction. Hence, by Stone's Theorem, one immediately has the spectral representation

$$\bar{\mathbf{P}}_t^{\mathcal{L}_\rho} = \int_{[0,\infty)} e^{-\lambda t}\,dE_\lambda,$$

where $\{E_\lambda : \lambda \in [0,\infty)\}$ is a resolution of the identity by orthogonal projections in $L^2(\nu_\rho;\mathbb{R})$. As this observation indicates, $\{\mathbf{P}_t^{\mathcal{L}_\rho} : t > 0\}$ enjoys the best of both the functional analytic as well as the probabilistic worlds. Thus, it should not be surprising that its analysis is particularly rewarding (cf. §7.5 in [35]).

3.6.3. An Application to the Explosion Problem. The purpose of this section is to extend Theorem 3.13 to cover cases in which the function u is not smooth. The sort of result for which we are looking says that the same conclusion holds when the function u in that theorem is suitably Lipschitz continuous[10] and satisfies $\alpha u - \mathcal{L}u \ge 0$ in the sense that (cf. (3.48)):

$$\int_{\mathbb{R}^N} (\mathcal{L}^*\varphi)u\,dx \le \alpha\int_{\mathbb{R}^N}\varphi u\,dx \quad \text{for all } \varphi \in C_c^\infty(\mathbb{R}^N;[0,\infty)).$$

Although the proof which we give below can be easily modified (cf. [37]) to cover the general case, we will restrict our attention to the case when $\mathcal{L} = \mathcal{L}_\rho$, where $\mathcal{L}_\rho$ is given by (3.56).

3.60 LEMMA. *Let W be a bounded open subset of $\mathbb{R}^N$, $f \in C(W;[0,\infty))$, and assume that $u \in C(W;[0,\infty))$ satisfies the conditions*

$$\max_{1\le k\le d}\left|\int_W (X_k^*\varphi)\,u\,dx\right| \le C\int_W |\varphi|\,dx \tag{3.61}$$
$$\text{for some } C < \infty \text{ and all } \varphi \in C_c^\infty(W;\mathbb{R}),$$

[10] If one (cf. [37]) works a little harder, one can dispense with this assumption.

and

$$\int_W (\mathcal{L}_\rho\varphi)u\,d\nu_\rho \le \int_W \varphi(\alpha u + f)\,d\nu_\rho \tag{3.62}$$
$$\text{for some } \alpha > 0 \text{ and all } \varphi \in C_c^\infty(W;[0,\infty)).$$

Then, for any open U with $\overline{U} \subseteq W$, any $x \in U$, and any solution $\mathbb{P}$ to the martingale problem for $\mathcal{L}_\rho$ starting at x,

$$\mathbb{E}^{\mathbb{P}}\left[e^{-\alpha T}u(p(T\wedge\zeta^U)\right] \le u(x) + \mathbb{E}^{\mathbb{P}}\left[\int_0^{T\wedge\zeta^U} e^{-\alpha t} f(p(t))\,dt\right].$$

PROOF: Choose $\eta \in C_c^\infty(W;[0,1])$ so that $\eta \equiv 1$ in an open neighborhood of $\overline{U}$, set $\tilde{X}_k = \eta X_k$ for $1 \le k \le d$, and define $\tilde{\mathcal{L}}_\rho$ on $C^2(\mathbb{R}^N;\mathbb{R})$ by

$$\tilde{\mathcal{L}}_\rho f = -\frac{1}{2\rho}\sum_{k=1}^d \tilde{X}_k^*(\rho\tilde{X}_k f).$$

Then, by Theorem 3.10, $\mathbb{P} \restriction \mathcal{F}_{\zeta^U} = \mathbb{P}_x^{\tilde{\mathcal{L}}_\rho} \restriction \mathcal{F}_{\zeta^U}$; and so it suffices for us to check that

$$\mathbb{E}^{\mathbb{P}_x^{\tilde{\mathcal{L}}_\rho}}\left[e^{-\alpha T}u(p(T\wedge\zeta^U))\right] \le u(x) + \mathbb{E}^{\mathbb{P}_x^{\tilde{\mathcal{L}}}}\left[\int_0^{T\wedge\zeta^U} e^{-\alpha t} f(p(t))\,dt\right].$$

To this end, first observe that for any $\varphi \in C_c^\infty(W;[0,\infty))$:

$$2\hat{\mathcal{L}}_\rho\varphi = 2\mathcal{L}_\rho(\eta^2\varphi) + \frac{1}{\rho}\sum_{k=1}^d X_k^*(\rho(X_k\eta^2)\varphi).$$

Thus, by (3.61) and (3.62),

$$\int_W (\tilde{\mathcal{L}}_\rho\varphi)u\,d\nu_\rho \le \int_W \varphi v\,d\nu_\rho,$$

where $v \equiv \eta^2(\alpha u + f) + C\sum_{k=1}^d \eta|X_k\eta|$.

Next, because the $\tilde{X}_k$'s all vanish off a compact $K \subseteq W$, Theorem 2.68 together with the first part of Theorem 2.40 guarantee that

$$\mathbb{P}_x^{\tilde{\mathcal{L}}_\rho}\Big(p(t) \in \{x\} \cup K,\ t \in [0,\infty)\Big) = 1 \quad \text{for each } x \in \mathbb{R}^N. \tag{3.63}$$

Hence, if $\varphi \in C_c^\infty(W;[0,\infty))$ is given and we set $\varphi_t \equiv \mathbf{P}_t^{\tilde{\mathcal{L}}_\rho}\varphi$, then $t \rightsquigarrow \varphi_t$ is a smooth map from $[0,\infty)$ into $C_c^\infty(W;[0,\infty))$ and $\frac{d}{dt}\varphi_t = \hat{\mathcal{L}}_\rho\varphi_t$. Applying (3.59) and the preceding, we conclude that

$$\begin{aligned}\frac{d}{dt}\int_W \varphi\,\mathbf{P}_t^{\tilde{\mathcal{L}}_\rho}u\,d\nu_\rho &= \frac{d}{dt}\int_W \varphi_t\,u\,d\nu_\rho \\ &= \int_W (\tilde{\mathcal{L}}_\rho\varphi_t)u\,d\nu_\rho \le \int_W \varphi_t\,v d\nu_\rho = \int_W \varphi\mathbf{P}_t^{\tilde{\mathcal{L}}_\rho}v\,d\nu_\rho.\end{aligned}$$

After integrating with respect to t and using the fact that both u and v are continuous, we arrive at

$$\mathbf{P}_t^{\bar{\mathcal{L}}_\rho} u \le u(x) + \int_0^t \mathbf{P}_\tau^{\bar{\mathcal{L}}_\rho} v \, d\tau, \quad (t,x) \in [0,\infty) \times W.$$

Alternatively, by the Markov property and (3.63),

$$\mathbb{E}^{\mathbb{P}_x^{\bar{\mathcal{L}}_\rho}}\left[u\big(p(s+t)\big) \,\middle|\, \mathcal{F}_s\right] = \left[\mathbf{P}_t^{\bar{\mathcal{L}}_\rho} u\right]\big(p(s)\big)$$
$$\le u\big(p(s)\big) + \int_0^t \left[\mathbf{P}_\tau^{\bar{\mathcal{L}}_\rho} v\right]\big(p(s)\big)\, d\tau = u\big(p(s)\big) + \mathbb{E}^{\mathbb{P}_x^{\bar{\mathcal{L}}_\rho}}\left[\int_s^{s+t} v\big(p(\tau)\big)\, d\tau \,\middle|\, \mathcal{F}_s\right],$$

and so

$$u\big(p(t)\big) - \int_0^t v\big(p(\tau)\big)\, d\tau$$

is a $\mathbb{P}_x^{\bar{\mathcal{L}}_\rho}$-supermartingale for each $x \in W$. But, by Doob's Stopping Time Theorem, this means that

$$F(t) \equiv \mathbb{E}^{\mathbb{P}_x^{\bar{\mathcal{L}}_\rho}}\left[u\big(p(t \wedge \zeta^U)\big)\right] \le u(x) + \mathbb{E}^{\mathbb{P}_x^{\bar{\mathcal{L}}_\rho}}\left[\int_0^{t\wedge\zeta^U} v\big(p(\tau)\big)\, d\tau\right]$$
$$\le u(x) + \alpha \int_0^t F(\tau)\, d\tau + \int_0^t \mathbb{E}^{\mathbb{P}}\left[f\big(p(\tau)\big),\, \zeta^U > \tau\right] d\tau,$$

since $v \restriction U = (\alpha u + f) \restriction U$ and $u \ge 0$. From here the desired estimate is an easy application of Gronwall's inequality. □

Given the conclusion drawn in Lemma 3.60, the following statement follows from the reasoning given in the last part of the proof of Theorem 3.13.

3.64 Theorem. *Suppose that* $(X_1, \dots, X_d)$ *are smooth vector fields on* $\mathbb{R}^N$, $\rho : \mathbb{R}^N \longrightarrow (0,\infty)$ *is smooth, and* $\mathcal{L}_\rho$ *is the associated operator given by (3.56). Further, assume* $u \in C\big(\mathbb{R}^N; [0,\infty)\big)$ *has the property that, for each bounded open* $W \subseteq \mathbb{R}^N$, *(3.61) holds with a* $C < \infty$, *which may depend on* W, *and*

$$\int_W (\mathcal{L}_\rho \varphi) u \, d\nu_\rho \le \alpha \int_W \varphi u \, d\nu_\rho, \quad \varphi \in C_c^\infty\big(W; [0,\infty)\big),$$

for some $\alpha \in (0,\infty)$ *which does not depend on* W. *If* u *has compact level sets, then the martingale problem for* $\mathcal{L}_\rho$ *is well-posed.*

CHAPTER 4

Doing it on a Manifold, An Extrinsic Approach

In this chapter we will show how the results obtained thus far can be literally *restricted* to a submanifold of $\mathbb{R}^N$. Of course, because the approach which I have in mind is unabashedly extrinsic, some readers may find it wanting in geometric insight. Nonetheless, I hope that even they will find it amusing.

4.1 Diffusions on a Submanifold of $\mathbb{R}^N$

Throughout this chapter, M will be a smooth, embedded, closed submanifold of $\mathbb{R}^N$. Thus, for each $x \in M$, we can identify the tangent space T_xM with the $\dim(M)$-dimensional subspace of $T_x\mathbb{R}^N$ consisting of those X_x which are *tangent to M at x* in the sense that there is a smooth path $p : \mathbb{R} \longrightarrow M \subseteq \mathbb{R}^N$ with the properties that $p(0) = x$ and $\dot{p}(0) = X_x$. Similarly, each X in the space of smooth vector fields on M can be thought of as the restriction to M of a smooth vector field $\tilde{X}$ on $\mathbb{R}^N$ with the property that $\tilde{X}_x = X_x \in T_xM$ for each $x \in M$. Of course, there are uncountably many $\tilde{X}$'s corresponding to the same vector field X on M. On the other hand, as long as one stays on M, there is no problem. In particular, if $x \in M$ and $p : [0, T) \longrightarrow \mathbb{R}^N$ is an integral curve of $\tilde{X}$ starting at x, then $p(t) \in M$ for each $t \in [0, T)$ and p is independent of which $\tilde{X}$ is chosen. Equivalently, for any $n \in \mathbb{N} \cup \{\infty\}$, we can realize each $f \in C^n(M; \mathbb{R})$ as the restriction to M of an $\tilde{f} \in C^n(\mathbb{R}^N; \mathbb{R})$ and, for $n \geq 1$, can evaluate X_xf as $\tilde{X}_x\tilde{f}$.

4.1.1. The Martingale Problem. Our goal is to prove that the preceding remarks about the relationship between integral curves on M and $\mathbb{R}^N$ have a precise analog in the context of martingale problems.

Suppose that $(X_0, \dots, X_d)$ are smooth vector fields on M, and define the operator $\mathcal{L}$ accordingly, as in (2.3). In view of the preceding remarks, it should be clear that, for any choice $(\tilde{X}_0, \dots, \tilde{X}_d)$ of smooth vector fields on $\mathbb{R}^N$ satisfying

$$\tilde{X}_k \restriction M = X_k, \quad 0 \leq k \leq d, \tag{4.1}$$

$\tilde{\mathcal{L}}\tilde{f} \restriction M = \mathcal{L}f$ for every $f \in C^2(M; \mathbb{R})$ and $\tilde{f} \in C^2(\mathbb{R}^N; \mathbb{R})$ with $\tilde{f} \restriction M = f$ when $\tilde{\mathcal{L}}$ is the operator corresponding to the $\tilde{X}_k$'s. Next, set $\mathcal{P}(M) =$

$C([0,\infty);M)$, endow $\mathcal{P}(M)$ with the topology of uniform convergence on compacts, and take $\mathcal{F}_t = \sigma(\{p(\tau) : \tau \in [0,t]\})$ for each $t \in [0,\infty)$. Just as before, we will say that $\mathbb{P} \in \mathbf{M}_1(\mathcal{P}(M))$ (the space of Borel probability measures on $\mathcal{P}(M)$) solves the martingale problem for $\mathcal{L}$ starting at $x \in M$ if (cf. (2.1)) $t \rightsquigarrow M_f^{\mathcal{L}}(t,p)$ is a $\mathbb{P}$-martingale relative to $\{\mathcal{F}_t : t \geq 0\}$ for each $f \in C_c^\infty(M;\mathbb{R})$. Similarly, we will say that the martingale problem for $\mathcal{L}$ is well-posed on M if the martingale problem for $\mathcal{L}$ has precisely one solution starting at each $x \in M$.

4.2 THEOREM. *Assume that M is a smooth, embedded, closed submanifold in $\mathbb{R}^N$, and refer to the preceding paragraph. Then $\tilde{\mathbb{P}}$ solves the martingale problem for $\tilde{\mathcal{L}}$ starting at $x \in M$ if and only if $\tilde{\mathbb{P}}(\mathcal{P}(M)) = 1$ and $\mathbb{P} = \tilde{\mathbb{P}} \restriction \mathcal{P}(M)$ solves the martingale problem for $\mathcal{L}$ starting at x. In particular, for each $x \in M$, there is at most one solution $\mathbb{P}_x^{\mathcal{L}}$ to the martingale problem for $\mathcal{L}$ starting at x, and one exists if and only if one does for $\tilde{\mathcal{L}}$, in which case $\mathbb{P}_x^{\mathcal{L}} = \mathbb{P}_x^{\tilde{\mathcal{L}}} \restriction \mathcal{P}(M)$. Finally, assume that X_{Ξ} (cf. (2.8)) is forward complete for every $\Xi \in [0,\infty) \times \mathbb{R}^d$, and, for each $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$ and $n \in \mathbb{N}$, determine (cf. §3.3.1) $p(\,\cdot\,,x,\mathbf{w}_n) \in \mathcal{P}(M)$ by (2.13), where $\mathbf{w}_n$ is the polygonal path obtained from $\mathbf{w}$ as in (3.17). Then $\mathbb{P}_x^{\mathcal{L}}$ exists if and only if there is a measurable $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto p(\,\cdot\,,x,\mathbf{w}) \in \mathcal{P}(M)$ such that*

$$p(\,\cdot\,,x,\mathbf{w}_n) \longrightarrow p(\,\cdot\,,x,\mathbf{w}) \quad \text{in } \mathcal{P}(M) \text{ for } \mu_{\mathbb{R}^d}\text{-almost every } \mathbf{w},$$

in which case $\mathbb{P}_x^{\mathcal{L}}$ is the $\mu_{\mathbb{R}^d}$-distribution of $\mathbf{w} \rightsquigarrow p(\,\cdot\,,x,\mathbf{w})$.

PROOF: Suppose that $\tilde{\mathbb{P}}$ solves the martingale problem for $\tilde{\mathcal{L}}$ starting at $x \in M$. Then, by Theorem 3.7, Corollary 3.11, and the preceding remark about integral curves, $\tilde{\mathbb{P}}(\mathcal{P}(M)) = 1$. Hence, since $\tilde{\mathcal{L}}\tilde{f} \restriction M = \mathcal{L}f$, it is trivial to check that $\tilde{\mathbb{P}} \restriction \mathcal{P}(M)$ solves the martingale problem for $\mathcal{L}$ starting at x. Conversely, if $\tilde{\mathbb{P}}(\mathcal{P}(M)) = 1$ and $\tilde{\mathbb{P}} \restriction \mathcal{P}(M)$ solves the martingale problem for $\mathcal{L}$ starting at $x \in M$, then it is trivial to see that $\tilde{\mathbb{P}}$ solves the martingale problem for $\tilde{\mathcal{L}}$ starting at x. Finally, to check the last part, simply apply the results in § 3.3 to the martingale problem for $\tilde{\mathcal{L}}$. □

Recall that if F is a differentiable map from M into itself, then, at each $x \in M$, F determines a linear map, the *pushforward map* $(F_*)_x$ from T_xM into $T_{F(x)}M$ by the prescription

$$(F_*)_x X_x f = X_x f \circ F, \quad f \in C^1(M;M). \tag{4.3}$$

Of course, in the standard coordinates for $\mathbb{R}^N$, $(F_*)_x$ is precisely the restriction to T_xM of the linear map on $\mathbb{R}^N$ whose matrix representation is the Jacobian matrix $\frac{\partial F}{\partial x}(x)$. Finally, the group $\mathrm{Diff}(M)$ of diffeomorphisms on M is the group of $F \in C^\infty(M;M)$ which are homeomorphic onto M with inverse also in $C^\infty(M;M)$; and we give $\mathrm{Diff}(M)$ the topology which makes $F \in \mathrm{Diff}(M) \longmapsto (F,F^{-1}) \in C^\infty(M;M)^2$ both open and continuous. By the Implicit Function Theorem, a homeomorphism $F \in C^\infty(M;M)$ is diffeomorphism if and only if

F_* is nowhere degenerate. In fact, if $\xi \rightsquigarrow F_\xi$ is a continuous map with values in $C^\infty(M;M)$ and, for each ξ, F_ξ is a homeomorphism onto M, then $\xi \rightsquigarrow F_\xi$ is a continuous map into $\mathrm{Diff}(M)$ if and only if, for each ξ, $(F_\xi)_*$ is nowhere degenerate.

The following corollary is a simple application of the first part of Theorem 4.2 and the results in § 3.5.

4.4 COROLLARY. *Assume that the martingale problem for $\mathcal{L}$ is well-posed on M. Then there is a measurable map*

$$\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto p(\,\cdot\,, *\,, \mathbf{w}) \in C^{0,\infty}\big([0,\infty) \times M; M\big)$$

such that:

(1) *for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$ (cf. (3.17) and Theorem 4.2)*

$$p(\,\cdot\,, *\,, \mathbf{w}_n) \longrightarrow p(\,\cdot\,, *\,, \mathbf{w}) \text{ in } C^{0,\infty}\big([0,\infty) \times M; M\big),$$

(2) *for each $x \in M$, the $\mu_{\mathbb{R}^d}$-distribution of $\mathbf{w} \rightsquigarrow p(\,\cdot\,, x, \mathbf{w}) \in \mathcal{P}(M)$ is $\mathbb{P}^{\mathcal{L}}_x$,*
(3) *for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$ and every $t \in [0,\infty)$, $p(t, *\,, \mathbf{w})$ is one-to-one and $p(t, *\,, \mathbf{w})_*$ is nowhere degenerate.*

In particular, $x \in M \longmapsto \mathbb{P}^{\mathcal{L}}_x \in \mathbf{M}_1\big(\mathcal{P}(M)\big)$ is weakly continuous and (3.16) holds for every $\{\mathcal{F}_t : t \geq 0\}$ stopping time $\zeta : \mathcal{P}(M) \longrightarrow [0,\infty]$. Finally, if $\check{L}$ is given by (3.45) and the martingale problem for $\check{L}$ is also well-posed on M, then, for each $t \in [0,\infty)$, $p(t, *\,, \mathbf{w}) \in \mathrm{Diff}(M;M)$ for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$.

4.1.2. Invariant and Symmetric Measures. Our aim here is to transfer the results in §3.6 to the present setting, and most of the problem is to figure out the appropriate differential geometric meaning of the quantities which enter the statements of Theorems 3.40 and 3.50.

For one thing, we need to put a standard reference measure λ_M on M to play the role which Lebesgue measure plays on $\mathbb{R}^N$, and for our purposes the best choice is the one determined by the Riemannian structure which M inherits as a submanifold of $\mathbb{R}^N$. Namely, given a coordinate chart (U,Φ) for M containing x, define the Borel measure $\lambda_{(U,\Phi)}$ on U so that

$$\int_U f(y)\,\lambda_{(U,\Phi)}(dy) = \int_{\Phi(U)} f \circ \Phi^{-1}(w)\sqrt{\det\big(g^\Phi \circ \Phi^{-1}(w)\big)}\,dw \tag{4.5}$$

for all $f \in C\big(U;[0,\infty)\big)$, where

$$g^\Phi(x) \equiv \Big(\Big(\big((\partial^\Phi_m)_x, (\partial^\Phi_{m'})_x\big)_{\mathbb{R}^N}\Big)\Big)_{1 \leq m,m' \leq \dim(M)} \tag{4.6}$$

and

$$\partial^\Phi_m \equiv (\Phi^{-1})_*\partial_m \tag{4.7}$$

is the vector field on U determined by

$$\Phi_*^{-1}\partial_m f = \partial_m\big(f \circ \Phi^{-1}\big).$$

Since if (U', Φ') is a second chart, then

$$g^{\Phi}(x) = \left(\frac{\partial \Phi' \circ \Phi^{-1}}{\partial u}\big(\Phi(x)\big)\right)^{\top} g^{\Phi}(x) \left(\frac{\partial \Phi' \circ \Phi^{-1}}{\partial u}\big(\Phi(x)\big)\right) \tag{4.8}$$

for $x \in U \cap U'$, an application of Jacobi's change of variable formula shows that the restrictions of $\lambda_{(U,\Phi)}$ and $\lambda_{(U',\Phi')}$ to $U \cap U'$ coincide. Hence, we can determine a unique Borel measure λ_M on M by insisting that $\lambda_M \restriction U = \lambda_{(U,\Phi)}$ for each chart (U, Φ). The measure λ_M is called the *Riemann measure* on the submanifold M .

Having constructed λ_M, we next want to find out what Jacobi's change of variable formula looks like for λ_M. To this end, let $F \in C^1(M; M)$ be given. If $\mathfrak{b}(x) = \big((B_1)_x, \dots, (B_{\dim(M)})_x\big)$ is a basis in T_xM, set

$$JF\big(x, \mathfrak{b}(x)\big) = \Big(\Big(\big((B_k)_x F, (B_\ell)_x F\big)_{\mathbb{R}^N}\Big)\Big)_{1 \le k,\ell \le \dim(M)}.$$

In particular, if I_M denotes the restriction of the identity map to M, then $JI_M\big(x, \mathfrak{b}(x)\big)$ is non-degenerate and, by elementary linear algebra, the ratio of $\det\big(J(F(x, \mathfrak{b}(x))\big)$ to $\det\big(JI_M(x, \mathfrak{b}(x))\big)$ is independent of the choice of basis $\mathfrak{b}(x)$. Hence, we can define the *Jacobian* jF of F at x by

$$jF(x) = \sqrt{\frac{\det\big(JF(x, \mathfrak{b}(x))\big)}{\det\big(JI_M(x, \mathfrak{b}(x))\big)}}. \tag{4.9}$$

Clearly, if $F \in C^{n+1}(M; M)$, then $(jF)^2 \in C^n(M; M)$. In addition, $F \rightsquigarrow jF$ is multiplicative in the sense that

$$F, G \in C^1(M; M) \implies j(F \circ G) = \big((jF) \circ G\big) jG.$$

In particular, if $F \in \mathrm{Diff}(M)$, then

$$jF^{-1} = \frac{1}{(jF) \circ F^{-1}}. \tag{4.10}$$

4.11 THEOREM. *If $F \in \mathrm{Diff}(M)$, then*

$$\int_M f \circ F(x)\, \lambda_M(dx) = \int_M f(y)\, jF^{-1}(y)\, \lambda_M(dy) \quad \text{for all } f \in C\big(M; [0, \infty)\big)$$

PROOF: It suffices to handle f's which are compactly supported in some coordinate chart (U',Φ'). Given such an f, set $U = F^{-1}(U')$ and $\Phi = \Phi' \circ F$. Then (U,Φ) is a coordinate chart, and so, by (4.5),

$$\begin{aligned}\int_U f\circ F(x)\,\lambda_M(dx) &= \int_{\Phi(U)} f\circ F\circ\Phi^{-1}(w')\sqrt{\det\bigl(g^\Phi\circ\Phi^{-1}(w')\bigr)}\,dw'\\ &= \int_{\Phi'(U')} f\circ(\Phi')^{-1}(w')\sqrt{\det\bigl(g^\Phi\circ F^{-1}\bigr)\circ(\Phi')^{-1}(w')}\,dw'\\ &= \int_{U'} f(y)\sqrt{\frac{\det\bigl(g^\Phi\circ F^{-1}(y)\bigr)}{\det\bigl(g^{\Phi'}(y)\bigr)}}\,\lambda_M(dy).\end{aligned}$$

Finally, observe that

$$\frac{\det\bigl(g^\Phi\circ F^{-1}(y)\bigr)}{\det\bigl(g^{\Phi'}(y)\bigr)} = \frac{\det\bigl(JF^{-1}(y,\mathfrak{b}(y))\bigr)}{\det\bigl(JI_M(y,\mathfrak{b}(y))\bigr)} = \bigl(jF^{-1}(y)\bigr)^2,$$

when $\mathfrak{b}(y) = \bigl((\partial_1^{\Phi'})_y,\dots,(\partial_{\dim(M)}^{\Phi'})_y\bigr)$ is the basis in T_yM determined from the coordinate system Φ'. $\square$

The final ingredients which we need are the divergence of a vector field and its relation to the flow determined by that vector field. Thus, let X be a vector field on M. Given a chart (U,Φ) and an $f\in C_c^\infty(U;\mathbb{R})$, it is clear from (4.5) that

$$(4.12)\qquad \int_M Xf\,d\lambda_M = -\int_M f\,\mathrm{div}^M(X)\,d\lambda_M,$$

where

$$(4.13)\qquad \mathrm{div}^M(X) \equiv \frac{1}{\sqrt{\det(g^\Phi)}}\sum_{m=1}^{\dim(M)}\partial_m^\Phi\left(\sqrt{\det(g^\Phi)}\,X\Phi^m\right).$$

Although the right hand side of (4.13) appears to depend on the choice of coordinate system, the equation (4.12) shows that the formal λ_M-adjoint of X (as a first order differential operator) is given by

$$(4.14)\qquad X^{*\lambda_M} = -X - \mathrm{div}^M(X),$$

which proves that the expression in (4.13) for $\mathrm{div}^M(X)$ must be independent of the choice of coordinate system.

Next, assume that X is a complete vector field on M, and, for each $x\in M$, let $p_X(\,\cdot\,,x)\in C^\infty(\mathbb{R};M)$ be the integral curve of X starting at x. Then, for each $s\in\mathbb{R}$, $p_X(s,\,\cdot\,)\in\mathrm{Diff}(M)$ and $s\in\mathbb{R}\longmapsto p_X(s,\,\cdot\,)\in\mathrm{Diff}(M)$ is a flow, in the sense that $p_X(\sigma+s,x) = p_X(s,p_X(\sigma,x))$. Our goal is to check that

$$(4.15)\qquad \bigl[jp_X(s,\,\cdot\,)\bigr](x) = \exp\left(\int_0^s\bigl[\mathrm{div}^M(X)\bigr]\bigl(p_X(\sigma,x)\bigr)\,d\sigma\right).$$

To this end, note that, for any pair $\varphi, \psi \in C_c^\infty(M;\mathbb{R})$,

$$\begin{aligned}\int_M \psi X^{*\lambda_M}\varphi\, d\lambda_M &= \frac{d}{ds}\int_M \psi\big(p_X(s,x)\big)\,\varphi(x)\lambda_M(dx)\bigg|_{s=0}\\ &= \frac{d}{ds}\int_M \psi(x)\,\varphi\big(p_X(-s,x)\big)\big[jp_X(-s,\,\cdot\,)\big](x)\,\lambda_M(dx)\bigg|_{s=0},\end{aligned}$$

and therefore

$$\big[X^{*\lambda_M}\varphi\big](x) = \frac{d}{ds}\varphi\big(p_X(-s,x)\big)\big[jp_X(-s,\,\cdot\,)\big](x)\bigg|_{s=0}.$$

Next note that, by the flow property, $p_X(-s+\sigma,x) = p_X\big(-s, p_X(\sigma,x)\big)$ and

$$\big[jp_X(-s+\sigma,\,\cdot\,)\big](x) = \big[jp_X(-s,\,\cdot\,)\big]\big(p_X(\sigma,x)\big)\big[jp_X(\sigma,\,\cdot\,)\big](x).$$

Hence, after replacing x by $p_X(\sigma,x)$ in the preceding, we find that

$$(4.16)\quad \big[X^{*\lambda_M}\varphi\big]\big(p_X(\sigma,x)\big)\big[jp_X(\sigma,\,\cdot\,)\big](x) = -\frac{d}{d\sigma}\Big(\varphi\big(p_X(\sigma,x)\big)\big[jp_X(\sigma,\,\cdot\,)\big](x)\Big).$$

Knowing that (4.16) holds for all $\varphi \in C_c^\infty(M;\mathbb{R})$ and using (4.14), one quickly concludes first that

$$\frac{d}{ds}jp_X(s,\,\cdot\,) = \big[\mathrm{div}(X)\big]\big(p_X(s,\,\cdot\,)\big)jp_X(s,\,\cdot\,)$$

and thence that (4.15) holds.

It should be now clear that we have all the information that we need to reproduce for closed submanifolds of $\mathbb{R}^N$ the conclusions drawn in §§3.5.3, 3.6.1, and 3.6.2. That is, exactly the arguments used there can be repeated here to prove the results which are summarized in the following statement and its corollary.

4.17 THEOREM. *Let $(X_0,\dots,X_d)$ be smooth vector fields on the closed, embedded submanifold M of $\mathbb{R}^N$; and define the operators $\mathcal{L}$ and $\check{\mathcal{L}}$ on $C^2(M;\mathbb{R})$ accordingly, as in (2.3) and (3.45), respectively. Further, define the vector fields $(\hat{X}_1,\dots,\hat{X}_d)$ on $\mathbb{R}\times M$ by (cf. (3.49)) $\hat{X}_k = -\mathrm{div}^M(X_k)\partial_0 - X_k$, and take $\widehat{\mathcal{L}}$ to be the associated operator on $C^2(\mathbb{R}\times M;\mathbb{R})$. Finally, assume that the martingale problems for both $\mathcal{L}$ and $\check{\mathcal{L}}$ are well-posed on M. Then, the martingale problem for $\widehat{\mathcal{L}}$ is well-posed on $\mathbb{R}\times M$. Further, if $\{\mathbf{P}_t^{\mathcal{L}} : t>0\}$ and $\{\mathbf{P}_t^{\widehat{\mathcal{L}}} : t>0\}$ are the Feller continuous, Markov semigroups associated (cf. (3.46)) with these well-posed martingale problems, then, for any non-negative, continuous functions f and g on M,*

$$\int_M f(x)\big[\mathbf{P}_t^{\mathcal{L}}g\big](x)\,\lambda_M(dx) = \int_M f(x)\,\big[\mathbf{P}_t^{\widehat{\mathcal{L}}}\hat{g}\big](0,x)\,\lambda_M(dx),$$

where $\hat{g}(\xi, x) = e^{\xi} g(x)$ for $(\xi, x) \in \mathbb{R} \times M$. *Finally, if* $\mathcal{L}^{*\lambda_M}$ *denotes the formal* λ_M-*adjoint of* $\mathcal{L}$ and $\rho \in C^\infty(M; [0,\infty))$ *satisfies* $\mathcal{L}^{*\lambda_M}\rho \le 0$, *then* (3.54) *holds when* ν_ρ *is the Borel measure on* M *given by*

$$\nu_\rho(dx) = \rho(x)\,\lambda_M(dx). \tag{4.18}$$

In particular, if ν_ρ *is a probability measure, then*

$$\mathcal{L}^{*\lambda_M}\rho \le 0 \iff \nu_\rho = \nu_\rho \mathbf{P}_t^{\mathcal{L}} \iff \mathcal{L}^{*\lambda_M}\rho = 0.$$

4.19 COROLLARY. *Given* $\rho \in C^\infty(M; (0,\infty))$, *let* $\mathcal{L}_\rho$ *be the operator on* $C^2(M;\mathbb{R})$ *given by (cf. (4.14))*

$$\mathcal{L}_\rho f = -\frac{1}{\rho}\sum_{k=1}^{d} X_k^{*\lambda_M}(\rho X_k f). \tag{4.20}$$

If the martingale problem for $\mathcal{L}_\rho$ *is well-posed, then* (3.59) *holds for all non-negative, continuous* f *and* g *on* M.

4.1.3. Non-Explosion Criterion. Without changing the argument in any way, we can proceed as in the proof of Theorem 3.64 to pass from the results in the preceding subsection to the following non-explosion criterion for diffusions on M.

4.21 THEOREM. *Let* $(X_1, \dots, X_d)$, ρ, and $\mathcal{L}_\rho$ *be as in the Corollary 4.19,* and *let* u *be a non-negative, continuous function with compact level sets on the closed submanifold* M. *Further, assume that, for each bounded open* $W \subseteq M$, *there is a* $C_W < \infty$ *such that*

$$\left|\int_W (X^{*\lambda_M}\varphi)\, u\, d\lambda_M\right| \le C_W \int_W |\varphi|\, d\lambda_M \quad \text{for all } \varphi \in C_c^\infty(W;\mathbb{R}).$$

If, in addition, there is a $\alpha > 0$ *such that*

$$\int_M (\mathcal{L}_\rho \varphi)\, u\, d\lambda_M \le \alpha \int_M \varphi\, u\, d\lambda_M \quad \text{for all } \varphi \in C_c^\infty(M;[0,\infty)),$$

then the martingale problem for $\mathcal{L}_\rho$ *on* M *is well-posed.*

4.2 Brownian Motion on a Submanifold

Again, M is a smooth, embedded, closed submanifold of $\mathbb{R}^N$. Among the second order elliptic differential operators on $C^\infty(M;\mathbb{R})$ which one might want to consider, the one with the greatest geometric promise is the Laplacian Δ_M.

There is a myriad of ways in which to describe Δ_M, but the one which bears the greatest formal resemblance to a familiar expression for standard Laplacian $\Delta_{\mathbb{R}^N}$ is

$$\Delta_M f = \operatorname{div}^M(\operatorname{grad}^M f), \tag{4.22}$$

where, for $f \in C^\infty(M;\mathbb{R})$, the *gradient* $\mathrm{grad}^M f$ is the vector field on M determined by

$$(4.23)\qquad \big(X_x, \mathrm{grad}^M_x f\big)_{\mathbb{R}^N} = X_x f \quad \text{for } x \in M \text{ and } X_x \in T_x M,$$

and we have used $\mathrm{grad}^M_x f$ to denote the value $(\mathrm{grad}^M f)_x$ of $\mathrm{grad}^M f$ at x. Notice that if (U,Φ) is a coordinate chart, then (cf. (4.6) and (4.7))

$$(4.24)\qquad \mathrm{grad}^M f = \sum_{m=1}^{\dim(M)} \left(\sum_{m'=1}^{\dim(M)} \left(g^\Phi\right)^{m,m'} \partial^\Phi_{m'} f \right) \partial^\Phi_m$$

on U, where we have used the *raised indices* on g^Φ to indicate we are dealing with the *inverse* matrix. Hence (cf. (4.13)),

$$(4.25)\qquad \Delta_M f = \frac{1}{\sqrt{\det(g^\Phi)}} \sum_{m,m'=1}^{\dim(M)} \partial^\Phi_m \left(\sqrt{\det(g^\Phi)}\, \left(g^\Phi\right)^{m,m'} \partial_{m'} f \right)$$

on U.

One of the major virtues of the expression in (4.22) is that, together with (4.14), it leads immediately to

$$(4.26)\quad -\int_M g \Delta_M f \, d\lambda_M = \int_M \big(\mathrm{grad}^M f, \mathrm{grad}^M g\big)_{\mathbb{R}^N} \, d\lambda_M = -\int_M f \Delta_M g \, d\lambda_M,$$

for all $f,\, g \in C^\infty_c(M;\mathbb{R})$. Indeed, the first equality is just an application of (4.12) with $X = \mathrm{grad}^M f$, and the second equality comes from the first when the roles of f and g are reversed. In particular, this proves that $\Delta_M \restriction C^\infty_c(M;\mathbb{R})$ is symmetric in $L^2(M;\mathbb{R})$. On the other hand, neither the expression in (4.22) nor the one in (4.25) lends itself particularly well to the analysis which we have been doing. In particular, neither of these makes it clear that Δ_M can be written in the Hörmander form on the right hand side of (2.3). Of course, because g^Φ is non-degenerate, one could always work locally, as suggested at the beginning of § 2.1, to write Δ_M in terms of vector fields on each chart and then use a partition of unity to patch these local representations into a global one. However, the end result would be a rather *ad hoc* and not very satisfying. In fact, if M is not compact, it is not clear that this approach would allow a construction that does not require infinitely many vector fields. For this reason, we will spend the next section developing a more "natural" way to represent Δ_M in Hörmander form. Of course, what is "natural" depends on the eye of the beholder. In particular, the expression which we are about to develop can be "natural" only to someone who does not identify the term "natural" with the term "intrinsic."

4.2.1. Extrinsic Expressions. Given an orthonormal basis $\mathfrak{e} = (\mathbf{e}_1, \dots, \mathbf{e}_N)$ for $\mathbb{R}^N$, set

$$(\partial_i^{\mathfrak{e}})_x f = \frac{d}{d\xi} f(x + \xi \mathbf{e}_i)\big|_{\xi=0} \quad \text{for } 1 \le i \le N \text{ and } x \in \mathbb{R}^N. \tag{4.27}$$

Next, define the vector field $D_i^{\mathfrak{e},M}$ on M so that[1]

$$\left(D_i^{\mathfrak{e},M}\right)_x = \Pi_x^M (\partial_i^{\mathfrak{e}})_x, \quad 1 \le i \le N \text{ and } x \in M, \tag{4.28}$$

where Π_x^M is orthogonal projection operator from $T_x\mathbb{R}^N$ onto T_xM. It is easy to check that each $D_i^{\mathfrak{e},M}$ is a smooth vector field on M. In addition, it is clear that

$$\begin{aligned} X_x &= \sum_{i=1}^N \left(X_x, (\partial_i^{\mathfrak{e}})_x\right)_{\mathbb{R}^N} (D_i^{\mathfrak{e},M})_x = \sum_{i=1}^N \left(X_x, (D_i^{\mathfrak{e},M})_x\right)_{\mathbb{R}^N} (D_i^{\mathfrak{e},M})_x \\ &= \sum_{i=1}^N \left(X_x, (D_i^{\mathfrak{e},M})_x\right)_{\mathbb{R}^N} \partial_i^{\mathfrak{e}} \quad \text{for all } x \in M \text{ and } X_x \in T_xM. \end{aligned} \tag{4.29}$$

Thus, every vector field on M admits a global representation in terms of the $D_i^{\mathfrak{e},M}$'s. In particular, (4.29) leads to

$$\operatorname{grad}^M f = \sum_{i=1}^N (D_i^{\mathfrak{e},M} f) D_i^{\mathfrak{e},M} = \sum_{i=1}^N (D_i^{\mathfrak{e},M} f) \partial_i^{\mathfrak{e}}. \tag{4.30}$$

We next want to show that

$$\operatorname{div}^M(X) = \sum_{i=1}^N D_i^{\mathfrak{e},M} \left(X, D_i^{\mathfrak{e},M}\right)_{\mathbb{R}^N} = \sum_{i=1}^N D_i^{\mathfrak{e},M} \left(X, \partial_i^{\mathfrak{e}}\right)_{\mathbb{R}^N}, \tag{4.31}$$

and clearly it is only the first equality which needs any comment. To check this, first observe that the right hand side is independent of the choice of orthonormal basis $\mathfrak{e}$. Thus, given $x \in M$, it suffices to verify (4.31) at x for some choice of $\mathfrak{e}$. In particular, we can choose $\mathfrak{e} = (\mathbf{e}_1, \dots, \mathbf{e}_N)$ so that $(\mathbf{e}_1, \dots, \mathbf{e}_{\dim(M)})$ forms an orthonormal basis in T_xM and $\mathbf{e}_i \perp T_xM$ for $\dim(M) < i \le N$. Equivalently,

$$(D_i^{\mathfrak{e},M})_x = \begin{cases} (\partial_i^{\mathfrak{e},M})_x & \text{if } 1 \le i \le \dim(M) \\ \mathbf{0} & \text{if } \dim(M) + 1 \le i \le N. \end{cases} \tag{4.32}$$

[1] What follows is a rather serious abuse of notation. Namely, we are treating $\partial_i^{\mathfrak{e}}$ simultaneously as a symbol for partial differentiation and as an element of $\mathbb{R}^N$. In particular, a more correct expression for the quantity on the right of what follows would be $\partial_{\Pi_x^M \mathbf{e}_i}$.

At the same time, note that there exists a coordinate chart (U, Φ) with $x \in U$ and $\Phi(y) = (y^1, \dots, y^{\dim(M)})$ for $y \in U$, where $y^i = (y, \mathbf{e}_i)$. In particular, for each $1 \le m \le \dim(M)$,

$$\partial_m^\Phi = \partial_m^{\mathfrak{c}} + Y_m \quad \text{where}$$
$$(Y_m)_x = 0 \text{ and } (Y_m)_y \perp \operatorname{span}\bigl\{(\partial_1^{\mathfrak{c}})_y, \dots, (\partial_{\dim(M)}^{\mathfrak{c}})_y\bigr\} \text{ for all } y \in U.$$

Hence, we have that

$$\bigl\|g^\Phi(y) - I_{T_yM}\bigr\|_{\mathrm{H.S.}} = \mathcal{O}\bigl(|y - x|^2_{\mathbb{R}^N}\bigr),$$

and therefore that

$$\bigl[\operatorname{div}^M(X)\bigr](x) = \sum_{m=1}^{\dim(M)} (\partial_m^\Phi)_x\bigl(X, \partial_m^\Phi\bigr)_{\mathbb{R}^N} = \sum_{m=1}^{\dim(M)} (D_m^{\mathfrak{c},M})_x\bigl(X, \partial_m^\Phi\bigr)_{\mathbb{R}^N}.$$

At the same time,

$$\bigl(X_y, (\partial_m^\Phi)_y\bigr)_{\mathbb{R}^N} = \sum_{i=1}^N \bigl(X_y, (\partial_i^{\mathfrak{c}})_y\bigr)_{\mathbb{R}^N}\bigl((\partial_i^{\mathfrak{c}})_y, (\partial_m^\Phi)_y\bigr)_{\mathbb{R}^N},$$
$$\bigl(X_y, (\partial_i^{\mathfrak{c}})_y\bigr)_{\mathbb{R}^N} = O\bigl(|y - x|_{\mathbb{R}^N}\bigr) \quad \text{for } \dim(M) + 1 \le i \le N,$$

and

$$\bigl((\partial_i^{\mathfrak{c}})_y, (\partial_m^\Phi)_y\bigr)_{\mathbb{R}^N} = \begin{cases} \delta_{i,m} & \text{if } 1 \le i \le \dim(M) \\ O\bigl(|y - x|_{\mathbb{R}^N}\bigr) & \text{if } \dim(M) + 1 \le i \le N. \end{cases}$$

Thus,

$$\bigl[\operatorname{div}^M(X)\bigr](x) = \sum_{m=1}^{\dim(M)} (D_m^{\mathfrak{c},M})_x\bigl(X, \partial_m^{\mathfrak{c}}\bigr)_{\mathbb{R}^N} = \sum_{i=1}^N (D_i^{\mathfrak{c},M})_x\bigl(X, \partial_i^{\mathfrak{c}}\bigr)_{\mathbb{R}^N},$$

and so the first equality in (4.31) has now been proved.

By combining (4.22), (4.30), and (4.31), we obtain

$$\Delta_M f = \sum_{i=1}^N (D_i^{\mathfrak{c},M})^2 f, \tag{4.33}$$

which was the primary goal of this section. Notice that a dividend of (4.33), (4.26), (4.30), and (4.14) is the observation that

$$\begin{aligned} -\int_M g\Delta_M f \, d\lambda_M &= \sum_{i=1}^N \int_M D_i^{\mathfrak{c},M} g \, D_i^{\mathfrak{c},M} f \, d\lambda_M \\ &= \sum_{i=1}^N \int_M g \, (D_i^{\mathfrak{c},M})^{*\lambda_M} \circ D_i^{\mathfrak{c},M} f \, d\lambda_M \end{aligned}$$

for any pair $f, g \in C_c^\infty(M;\mathbb{R})$. Hence,

$$\Delta_M = -\sum_{i=1}^{N} (D_i^{\mathfrak{e},M})^{*\lambda_M} \circ D_i^{\mathfrak{e},M}. \tag{4.34}$$

4.2.2. Extrinsic Brownian Motion. Let M be a closed, embedded, submanifold of $\mathbb{R}^N$. If nothing else, (4.33) allows us to discuss the martingale problem for $\frac{1}{2}\Delta_M$. Indeed, all the results of § 4.1 apply when we take $d = N$, $X_0 \equiv \mathbf{0}$, and $X_k = D_k^{\mathfrak{e},M}$ for $1 \le k \le N$, in which case (4.33) says that $\frac{1}{2}\Delta_M$ is the operator $\mathcal{L}$ in (2.3) determined by these X_k's. In particular, we know that *there is at most one solution $\mathbb{P}_x^M \equiv \mathbb{P}_x^{\frac{1}{2}\Delta_M}$ to the martingale problem for $\frac{1}{2}\Delta_M$ starting at each $x \in M$ and that one exists unless there is explosion.* For historical reasons, we will say that a random variable $\omega \in \Omega \longmapsto p(\,\cdot\,,\omega) \in \mathcal{P}(M)$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a $\mathbb{P}$-*Brownian motion on M* starting at $x \in M$ if the $\mathbb{P}$-distribution of $\omega \rightsquigarrow p(\,\cdot\,,\omega)$ is

$$\mathbb{P}_x^M \equiv \mathbb{P}_x^{\frac{1}{2}\Delta_M}. \tag{4.35}$$

Actually, (4.33) allows us to say much more. In fact, it gives us a very pleasing geometric picture of the way one can go about massaging Brownian motion on $\mathbb{R}^N$ into Brownian motion on M. Namely, let $\mathfrak{e}$ be an orthonormal basis for $\mathbb{R}^N$. When $d = N$, $X_0 \equiv \mathbf{0}$, $X_k = D_k^{\mathfrak{e},M}$ for $1 \le k \le N$, and $\mathbf{w}$ is replaced by (cf. (3.17)) $\mathbf{w}_n$, (2.13) becomes

$$\dot p(t, x, \mathbf{w}_n) = \Pi^M_{p(t,x,\mathbf{w}_n)} \dot{\mathbf{w}}_n(t) \quad t \in (0,\infty) \text{ with } p(0, x, \mathbf{w}_n) = x. \tag{4.36}$$

Now Theorem 4.2 together with Corollary 3.58 and (4.34) give on the following statement.

4.37 Theorem. *Let $x \in M$ be given. Then a solution $\mathbb{P}_x^M$ to the martingale problem for $\frac{1}{2}\Delta_M$ starting at x exists if and only if there is a measurable $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^N) \longmapsto p(\,\cdot\,, x, \mathbf{w}) \in \mathcal{P}(M)$ such that*

$$p(\,\cdot\,, x, \mathbf{w}_n) \longrightarrow p(\,\cdot\,, x, \mathbf{w}) \quad \text{in } \mathcal{P}(M) \;\; (\text{a.s.}, \mu_{\mathbb{R}^N}),$$

where $p(\,\cdot\,, x, \mathbf{w}_n)$ is the solution to (4.36). Moreover, if $\mathbb{P}_x^M$ exists, then $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^N) \longmapsto p(\,\cdot\,, x, \mathbf{w}) \in \mathcal{P}(\mathbb{R}^N)$ is a $\mu_{\mathbb{R}^N}$-Brownian motion on M starting at x. In addition, if the martingale problem for $\frac{1}{2}\Delta_M$ is well-posed on M, then there exists a measurable

$$\mathbf{w} \in \mathfrak{W}(\mathbb{R}^N) \longmapsto p(\,\cdot\,, *, \mathbf{w}) \in C^{0,\infty}([0,\infty) \times M; M)$$

such that, for $\mu_{\mathbb{R}^N}$-almost every $\mathbf{w}$,

$$p(\,\cdot\,, *, \mathbf{w}_n) \longrightarrow p(\,\cdot\,, *, \mathbf{w}) \quad \text{in } C^{0,\infty}([0,\infty) \times M; M)$$

and, for each $t \in [0, \infty)$, $p(t, *, \mathbf{w}) \in \mathrm{Diff}(M)$. *In particular,* $x \in M \longmapsto \mathbb{P}_x^M \in \mathbf{M}_1(\mathcal{P}(M))$ *is continuous and (cf. (3.16))*

$$\mathbb{P}_x^M = \mathbb{P}_x^M \underset{\zeta}{\otimes} \mathbb{P}_{\cdot}^M$$

for every $\{\mathcal{F}_t : t \geq 0\}$*-stopping time* $\zeta : \mathcal{P}(M) \longrightarrow [0, \infty]$. *Finally, if, for* $f \in C(M; [0, \infty))$,

$$\left[\mathbf{P}_t^M f\right](x) = \mathbb{E}^{\mathbb{P}_x^M}\left[f\left(p(t)\right)\right], \quad (t, x) \in (0, \infty) \times M,$$

then $\{\mathbf{P}_t^M : t > 0\}$ *determines a Markov semigroup on* $C_b(M; \mathbb{R})$, *and*

$$\int_M g\, \mathbf{P}_t^M f\, d\lambda_M = \int_M f\, \mathbf{P}_t^M g\, d\lambda_M$$

for all $t > 0$ *and* $f, g \in C(M; [0, \infty))$.

It is important to remark at that, in some sense, the relationship between Brownian motion on $\mathbb{R}^N$ and Brownian motion of a submanifold M is simpler than that between geodesics on $\mathbb{R}^N$ and geodesics on M. Indeed, it is impossible to transform geodesics (i.e., straight lines) in $\mathbb{R}^N$ into geodesics on M by simple projection. The reason why projection works for Brownian motion is that $\mathbb{R}^N$-valued Brownian paths "spurge" in all directions simultaneously whereas a straight line has a well-defined direction and, as the geodesic in M turns away from that direction, the velocity of the straight line simply cannot keep up.

4.2.3. Brownian Motion Normal to a Submanifold. In connection with the discussion at the end of the preceding subsection, it is interesting to look at the component of $x + \mathbf{w}$ which is *normal* to M. That is, consider

$$q(\,\cdot\,, x, \mathbf{w}) \equiv x + \mathbf{w} - p(\,\cdot\,, x, \mathbf{w}). \tag{4.38}$$

In the case when M is an affine subspace of $\mathbb{R}^N$, it is trivial to see that $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^N) \longmapsto p(\,\cdot\,, x, \mathbf{w}) - x \in \mathcal{P}(\mathbb{R}^N)$ is a $\dim(M)$-dimensional (linear) $\mu_{\mathbb{R}^N}$-Brownian motion, that $\mathbf{w} \rightsquigarrow q(\,\cdot\,, x, \mathbf{w})$ is an $(N - \dim(M))$-dimensional (linear) $\mu_{\mathbb{R}^N}$-Brownian motion which moves in the subspace which is normal to M, and that these two Brownian motions are independent of one another. Obviously, when M is not affine, it would be foolish to suppose that this independence persists. Nonetheless, there should still be some sort of inherent, strong orthogonality between $p(\,\cdot\,, x, \mathbf{w})$ and $q(\,\cdot\,, x, \mathbf{w})$, and it is this property that we want to examine more closely here. Specifically, our goal is to describe the conditional $\mu_{\mathbb{R}^N}$-distribution of $\mathbf{w} \rightsquigarrow q(\,\cdot\,, x, \mathbf{w})$ given the sub σ-algebra $\sigma(\{p(t, x, \mathbf{w}) : t \geq 0\})$.

In order to carry out our program, we begin by considering the $M \times \mathbb{R}^N$-valued process

$$R(\,\cdot\,, x, \mathbf{w}) = \left(p(\,\cdot\,, x, \mathbf{w}), q(\,\cdot\,, x, \mathbf{w})\right).$$

To describe the distribution of $R(\,\cdot\,, x, \mathbf{w})$, define the vector fields $(Z_1, \dots, Z_{2N})$ on $M \times \mathbb{R}^N$ so that

$$\begin{aligned}(Z_k)_{(x,y)} f &= (D_k^{\mathfrak{e},M})_x f(\,*\,, y)\\ (Z_{k+N})_{(x,y)} f &= \big(D_k^{\mathfrak{e},M}(x)\big)_y^{\perp} f(x, \,*\,)\end{aligned} \qquad \text{for } 1 \le k \le N,$$

where $x \rightsquigarrow (D_k^{\mathfrak{e},M}(x))^{\perp}$ maps $x \in M$ to the vector field $(D_k^{\mathfrak{e},M}(x))^{\perp}$ on $\mathbb{R}^N$ given at $y \in \mathbb{R}^N$ by

$$\big(D_k^{\mathfrak{e},M}(x)\big)_y^{\perp} f = (\Pi_x^M)^{\perp} (\partial_k^{\mathfrak{e}})_y f = \sum_{j=1}^N \big((\Pi_x^M)^{\perp} \mathbf{e}_k, \mathbf{e}_j\big)_{\mathbb{R}^N} (\partial_j^{\mathfrak{e}})_y f$$

and $(\Pi_x^M)^{\perp} = I - \Pi_x^M$ is orthogonal projection onto the perpendicular complement $(T_x M)^{\perp}$ of $T_x M$ in $\mathbb{R}^N$. Because, for each $n \in \mathbb{N}$, (cf. § 3.3)

$$\dot{R}(t, x, \mathbf{w}_n) = \sum_{k=1}^N \dot{w}_n^k(t) \big(Z_k + Z_{k+N}\big)_{R(t,x,\mathbf{w}_n)},$$

we see that the $\mu_{\mathbb{R}^N}$-distribution of $R(\,\cdot\,, x, \mathbf{w})$ solves the martingale problem for

$$\mathcal{L} \equiv \frac{1}{2} \sum_{k=1}^N \big(Z_k + Z_{k+N}\big)^2$$

starting at $(x, \mathbf{0})$. Although this observation already gives us an implicit description of everything that can be known about the $\mu_{\mathbb{R}^N}$-distribution of $\mathbf{w} \rightsquigarrow R(\,\cdot\,, x, \mathbf{w})$ (and therefore the conditional distribution which we are seeking), we are hoping that the strong orthogonality alluded to above will allow us to do better. Namely, we are going to use this orthogonality to re-write $\mathcal{L}$ in a form which is amenable to the results from § 3.4. In fact,

$$\begin{gathered}\mathcal{L} = Z_0 + \frac{1}{2} \sum_{k=1}^{2N} Z_k^2,\\ \text{where } (Z_0)_{(x,y)} \equiv -\frac{1}{2} \sum_{k=1}^N \big[\operatorname{div}^M (D_k^{\mathfrak{e},M})\big](x) \big(D_k^{\mathfrak{e},M}(x)\big)_y^{\perp}.\end{gathered} \tag{4.39}$$

To see this, first write

$$\big(Z_k + Z_{k+N}\big)^2 = Z_k^2 + Z_{k+N}^2 + 2 Z_{k+N} \circ Z_k + [Z_k, Z_{k+N}].$$

Second, observe that

$$\begin{aligned}\sum_{k=1}^N \big(Z_{k+N}\big)_{(x,y)} \circ Z_k &= \sum_{k=1}^N \sum_{i,j=1}^N \big(\mathbf{e}_i, (\Pi_x^M)^{\perp} \mathbf{e}_k\big)_{\mathbb{R}^N} \big(\mathbf{e}_j, \Pi_x^M \mathbf{e}_k\big)_{\mathbb{R}^N} (\partial_i^{\mathfrak{e}})_x \circ (\partial_j^{\mathfrak{e}})_y\\ &= \sum_{i,j=1}^N \big((\Pi_x^M)^{\perp} \mathbf{e}_i, \Pi_x^M \mathbf{e}_j\big)_{\mathbb{R}^N} (\partial_i^{\mathfrak{e}})_x \circ (\partial_j^{\mathfrak{e}})_y = \mathbf{0}.\end{aligned}$$

Finally, by (4.31),

$$\begin{aligned}\sum_{k=1}^{N}[Z_k, Z_{k+N}]_{(x,y)} &= -\sum_{k=1}^{N}\sum_{j=1}^{N}(D_k^{\mathfrak{e},M})_x\big(\mathbf{e}_j, \Pi^M \mathbf{e}_k\big)_{\mathbb{R}^N}(\partial_j^{\mathfrak{e}})_y \\ &= -\sum_{j=1}^{N}\left(\sum_{k=1}^{N}(D_k^{\mathfrak{e},M})_x\big(D_j^{\mathfrak{e},M}, D_k^{\mathfrak{e},M}\big)_{\mathbb{R}^N}\right)(\partial_j^{\mathfrak{e}})_y \\ &= -\sum_{j=1}^{N}\big[\operatorname{div}^M(D_j^{\mathfrak{e},M})\big](x)(\partial_j^{\mathfrak{e}})_y.\end{aligned}$$

But, after comparing (4.33) with (4.34), one sees that

$$\sum_{j=1}^{N}\operatorname{div}^M(D_j^{\mathfrak{e},M})\Pi^M\partial_j^{\mathfrak{e}} \equiv 0,$$

and therefore that

$$\sum_{j=1}^{N}\big[\operatorname{div}^M(D_j^{\mathfrak{e},M})\big](x)(\partial_j^{\mathfrak{e}})_y = \sum_{k=1}^{N}\big[\operatorname{div}^M(D_k^{\mathfrak{e},M})\big](x)\big(D_k^{\mathfrak{e},M}(x)\big)_y^{\perp}.$$

The importance of (4.39) is that it tells us that distribution of $\mathbf{w}\rightsquigarrow R(\,\cdot\,, x, \mathbf{w})$ under $\mu_{\mathbb{R}^N}$ coincides with the $\mu_{\mathbb{R}^{2N}}$-distribution of the processes

$$\breve{\mathbf{w}} \in \mathfrak{W}(\mathbb{R}^{2N}) \longmapsto \bar{R}(\,\cdot\,, (x, \mathbf{0}), \bar{\mathbf{w}}) \in \mathcal{P}(M \times \mathbb{R}^N)$$

which one constructs in terms of the vector fields $(Z_0, \dots, Z_{2N})$ on $\mathbb{R}^N \times \mathbb{R}^N$. In particular, by the Theorem 3.29, we now know that the conditional $\mu_{\mathbb{R}^N}$-distribution of $\mathbf{w}\rightsquigarrow q(\,\cdot\,, x, \mathbf{w})$ given $\sigma(\{p(t, x, \mathbf{w}) : t \geq 0\})$ is $\Gamma^{p(\,\cdot\,, x, \mathbf{w})}$, where, for $p \in \mathcal{P}(M)$, $\Gamma^p \in \mathbf{M}_1\big(\mathcal{P}(\mathbb{R}^N)\big)$ is the solution to the martingale problem starting at $\mathbf{0}$ for the time-dependent operator

$$(4.40)\quad t\rightsquigarrow\mathcal{L}_{t,p} \equiv \frac{1}{2}\sum_{k=1}^{N}\Big(D_k^{\mathfrak{e},M}\big(p(t)\big)^{\perp} - \big[\operatorname{div}^M(D_k^{\mathfrak{e},M})\big]\big(p(t)\big)\Big)\,D_k^{\mathfrak{e},M}\big(p(t)\big)^{\perp}.$$

Although this operator is time-dependent, it is translation invariant in space, and this fact makes it very easy to solve the equation

$$\frac{\partial u}{\partial t} + \mathcal{L}_{t,p}u = 0 \text{ in } [0, T) \times \mathbb{R}^N \quad \text{with } u(T, \,\cdot\,) = f.$$

Indeed, define the *mean curvature normal* vector $\mathbf{N}_x$ at $x \in M$ so that[2]

$$(4.41)\qquad \mathbf{N}_x = \sum_{i=1}^{N}\big[\operatorname{div}^M(D_i^{\mathfrak{e},M})\big](x)(\Pi_x^M)^{\perp}\mathbf{e}_i.$$

[2] This vector is a familiar object in the differential geometry of submanifolds. See, for example, page 34 in volume II of [24], where a general definition is given. The identification of the expression here with the one there will be made clearer by the discussion in § 5.2.3.

Then, for $\boldsymbol{\xi} \in \mathbb{R}^N$,

$$(t, y) \in [0, T) \times \mathbb{R}^N \longmapsto U_{\boldsymbol{\xi}}^T(t, y; p)$$
$$\equiv \exp\left[\sqrt{-1}\left(\boldsymbol{\xi}, y - \frac{1}{2}\int_t^T \mathbf{N}_{p(\tau)}\, d\tau\right)_{\mathbb{R}^N} + \frac{1}{2}\int_t^T \left|(\Pi_{p(\tau)}^M)^{\perp}\boldsymbol{\xi}\right|_{\mathbb{R}^N}^2 d\tau\right]$$

is the solution when $f(y) = \exp\left[\sqrt{-1}\,(\boldsymbol{\xi}, y)_{\mathbb{R}^N}\right]$. But this means that

$$\left(U_{\boldsymbol{\xi}}^T\left(t \wedge T, q(t \wedge T); p\right), \mathcal{F}_t, \Gamma^p\right)$$

is a martingale, and so we find that

$$\mathbb{E}^{\Gamma^p}\left[\exp\left(\sqrt{-1}\,(\boldsymbol{\xi}, q(T) - q(t))_{\mathbb{R}^N}\right) \middle| \mathcal{F}_t\right]$$
$$= \exp\left[-\sqrt{-1}\left(\boldsymbol{\xi}, \frac{1}{2}\int_t^T \mathbf{N}_{p(\tau)}\, d\tau\right)_{\mathbb{R}^N} - \frac{1}{2}\int_t^T \left|(\Pi_{p(\tau)}^M)^{\perp}\boldsymbol{\xi}\right|_{\mathbb{R}^N}^2 d\tau\right]$$

for $0 \le t \le T$. That is, under Γ^p, $q \rightsquigarrow q(T) - q(t)$ is independent of $\sigma(\{q(\tau) : \tau \in [0, t]\})$ and is an $\mathbb{R}^N$-valued Gaussian random variable with

$$\text{mean value } -\frac{1}{2}\int_t^T \mathbf{N}_{p(\tau)}\, d\tau \quad \text{and covariance } \int_t^T \left(\Pi_{p(\tau)}^M\right)^{\perp} d\tau.$$

In other words, we have now proved the following.

4.42 THEOREM. *Let $x \in M$ be given, assume that $\mathbb{P}_x^M$ exists, denote by $\mathbf{w} \rightsquigarrow p(\,\cdot\,, x, \mathbf{w})$ the process in Theorem 4.37, and take $\mathbf{w} \rightsquigarrow q(\,\cdot\,, x, \mathbf{w})$ accordingly, as in (4.38). Then, $\mu_{\mathbb{R}^N}$-almost surely, the conditional $\mu_{\mathbb{R}^N}$-distribution of $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^N) \longmapsto q(\,\cdot\,, x, \mathbf{w}) \in \mathcal{P}(\mathbb{R}^N)$ given $\sigma(\{p(\,\cdot\,, x, \mathbf{w}) : t \ge 0\})$ is that of a process with independent increments such that the increment $q(t, x, \mathbf{w}) - q(s, x, \mathbf{w})$ is a Gaussian random variable with*

$$\text{mean value } -\frac{1}{2}\int_s^t \mathbf{N}_{p(\tau)}\, d\tau \quad \text{and covariance } \int_s^t \left(\Pi_{p(\tau)}^M\right)^{\perp} d\tau.$$

4.2.4. An Extrinsic Non-Explosion Criterion for Brownian Motion. We conclude this section with a criterion which guarantees that the Brownian motion on M will never explode; equivalently, that the martingale problem for $\frac{1}{2}\Delta_M$ on M is well-posed.[3]

[3] It is not so easy to give explicit examples of closed, embedded submanifolds in $\mathbb{R}^N$ on which the Brownian motion explodes. In fact, the only reason why I know such examples must exist relies on J. Nash's renowned embedding theorem, which says that any complete, separable Riemannian manifold can be realized as a closed submanifold of $\mathbb{R}^N$ for some N. Thus, it suffices to produce an example of some abstract, complete, separable Riemannian manifold on which the Brownian motion explodes. Such examples are quite easy to construct and will be given in § 8.4.2.

4.43 THEOREM. *Let $x \in M \longmapsto \mathbf{N}_x \in (T_x M)^\perp$ be the mean curvature normal defined in (4.41). Then the martingale problem for $\frac{1}{2}\Delta_M$ on M is well-posed if there exists a $C < \infty$ such that*

$$\big(x, \mathbf{N}_x\big)_{\mathbb{R}^N} \le C\big(1 + |x|^2_{\mathbb{R}^N}\big), \quad x \in M.$$

PROOF: The proof is a trivial application of Theorem 3.13 once one observes that

$$\big[\Delta_M u\big](x) = \sum_{i,j=1}^N \big(\Pi_x^N \mathbf{e}_i, \mathbf{e}_j\big)_{\mathbb{R}^N} (\partial_i^c)_x \partial_j^c u + \sum_{i=1}^N \big(\mathbf{N}_x, \mathbf{e}_i\big)_{\mathbb{R}^N} (\partial_i^c)_x u.$$

Thus, when $u(x) = 1 + |x|^2$,

$$\big[\tfrac{1}{2}\Delta_M u\big](x) = \dim(M) + \tfrac{1}{2}\big(\mathbf{N}_x, x\big)_{\mathbb{R}^N}. \quad \square$$

Although we will, in a sense, improve on this criterion in § 5.3.2, the present result has virtues of its own. Namely, it proves that Brownian motion never explodes on any closed, embedded *minimal submanifold* of $\mathbb{R}^N$. Indeed minimality of a submanifold in $\mathbb{R}^N$ is precisely the condition that its mean curvature normal vanish identically. (See § 5.3.3 for more information.)

4.3 A Question of Measurable Interest [4]

Let M be a submanifold of $\mathbb{R}^N$, assume that the martingale problem for $\frac{1}{2}\Delta_M$ is well-posed on M, and define $\mathbf{w} \rightsquigarrow p(\,\cdot\,, x, \mathbf{w}_n)$ as in (4.36). Clearly, $p(\,\cdot\,, x, \mathbf{w}_n)$ is piecewise smooth, and, by Theorem 4.37, $\{p(\,\cdot\,, x, \mathbf{w}_n)\}_1^\infty$ is $\mu_{\mathbb{R}^N}$-almost surely convergent to a Brownian motion on M starting at x. Thus, we have a scheme for approximating Brownian motion on M be piecewise smooth paths. However, there is a serious technical objection to this scheme. Namely, the approximations are not measurable functions of the Brownian motion being approximated. In this section we will give one way to to overcome this objection.

4.3.1. An Internal Approximation Scheme. Given $p \in \mathcal{P}(M)$ and $n \in \mathbb{N}$, define $p_n \in \mathcal{P}(M)$ so that (cf. § 2.3.1)

$$\begin{aligned} p_n(0) = p(0) \text{ and } \dot{p}_n(t) &= 2^n \Pi^M_{p_n(t)} \Delta_{m,n} p \quad \text{for } t \in \big(T_{m,n}, T_{m+1,n}\big) \\ &\text{where } \Delta_{m,n} p \equiv p\big(T_{m+1,n}\big) - p\big(T_{m,n}\big). \end{aligned} \tag{4.44}$$

Clearly $p \in \mathcal{P}(M) \longmapsto p_n \in \mathcal{P}(M)$ is continuous and each p_n is piecewise smooth. What we want to show is that, for each $(T, x) \in (0, \infty) \times M$,

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \big|p_n(t) - p(t)\big|_{\mathbb{R}^N} = 0 \quad \mathbb{P}_x^M\text{-almost surely.} \tag{4.45}$$

[4] No essential use will be made of the material in this section. My advice is to skip it.

First observe that it suffices for us to prove (4.45) when M is compact. Indeed, if this is not already the case, then we can find a sequence $\{M_\ell\}_1^\infty$ of compact submanifolds of $\mathbb{R}^N$ which exhaust M in the sense that for each compact subset K of M there is an ℓ_K such that $K \subseteq M_\ell$ for all $\ell \geq \ell_K$. The general case can therefore be reduced to the compact one by consideration of the paths up to the time when they first leave $M \cap M_\ell$ and then letting $\ell \nearrow \infty$. Thus, throughout the discussion which follows, we will assume that M is compact.

The idea of our proof will be to compare these p_n's to the approximants given in (4.36). However, in order to do so, we will have to resurrect the part of the $\mathbb{R}^N$-Brownian path which is missing from p. For this purpose, take Γ^p to be the solution to the martingale problem for the (time dependent) operator in (4.40). That is, Γ^p is the inhomogeneous Gaussian process described at the end of Theorem 4.42. Next, define the probability $\mathbb{R}_x$ on $\mathcal{P}(M) \times \mathcal{P}(\mathbb{R}^N)$ so that the marginal distribution of $\mathbb{R}_x$ on $\mathcal{P}(M)$ is $\mathbb{P}_x^M$ and the conditional distribution of $q \in \mathcal{P}(\mathbb{R}^N)$ given $p \in \mathcal{P}(M)$ is Γ^p. Equivalently, if $A \in \mathcal{B}_{\mathcal{P}(M)}$ and $B \in \mathcal{B}_{\mathcal{P}(\mathbb{R}^N)}$, then

$$\mathbb{R}_x(A \times B) = \int_A \Gamma^p(B)\, \mathbb{P}_x^M(dp). \tag{4.46}$$

By Theorem 4.42, we know that $\mathbb{R}_x$ is the $\mu_{\mathbb{R}^N}$-distribution of (cf. (4.38))

$$\mathbf{w} \in \mathfrak{W}(\mathbb{R}^N) \longmapsto \big(p(\,\cdot\,, x, \mathbf{w}), q(\,\cdot\,, x, \mathbf{w})\big) \in \mathcal{P}(M) \times \mathcal{P}(\mathbb{R}^N).$$

Hence, if $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^N)$ is given by $\mathbf{w}(t) \equiv p(t) - p(0) + q(t)$ for $t \geq 0$, then the $\mathbb{R}_x$-distribution of $\mathbf{w}$ is $\mu_{\mathbb{R}^N}$ and, for each $(T, x) \in (0, \infty) \times M$, (cf. § 3.3.1)

$$\mathbb{E}^{\mathbb{R}_x}\left[\sup_{t \in [0,T]} \big|p(t, x, \mathbf{w}_n) - p(t)\big|_{\mathbb{R}^N}^2\right] \leq C(T) 2^{-\frac{n}{2}}, \tag{4.47}$$

where $p(\,\cdot\,, x, \mathbf{w}_n)$ is determined from $\mathbf{w}$ by (4.36).

For each $n \in \mathbb{N}$ and $m \geq 0$,

$$p_n(T_{m+1,n}) - p_n(T_{m,n}) = 2^n \left(\int_{I_{m,n}} \Pi_{p_n(t)}\, dt\right) \Delta_{m,n} p$$

$$p(T_{m+1,n}, x, \mathbf{w}_n) - p(T_{m,n}, x, \mathbf{w}_n) = 2^n \left(\int_{I_{m,n}} \Pi_{p(t,x,\mathbf{w}_n)}\, dt\right) \Delta_{m,n} \mathbf{w},$$

where $I_{m,n} \equiv [T_{m,n}, T_{m+1,n}]$, $\Delta_{m,n}\mathbf{w} = \mathbf{w}(T_{m+1,n}) - \mathbf{w}(T_{m,n})$, and $\Pi \equiv \Pi^M$.

Thus,

$$\begin{aligned}
&p_n(T_{m,n}) - p(T_{m,n}, x, \mathbf{w}_n) \\
&= \sum_{\ell=0}^{m-1} 2^n \left[\left(\int_{I_{\ell,n}} \Pi_{p_n(t)}\, dt \right) \Delta_{\ell,n} p - \left(\int_{I_{\ell,n}} \Pi_{p(t,x,\mathbf{w}_n)}\, dt \right) \Delta_{\ell,n}\mathbf{w} \right] \\
&= \sum_{\ell=0}^{m-1} 2^n \left(\int_{I_{\ell,n}} \left(\Pi_{p_n(t)} - \Pi_{p(t,x,\mathbf{w}_n)} \right) dt \right) \Delta_{\ell,n}\mathbf{w} \\
&\qquad - \sum_{\ell=0}^{m-1} 2^n \left(\int_{I_{\ell,n}} \Pi_{p_n(t)}\, dt \right) \Delta_{\ell,n} q \quad \text{where } \Delta_{\ell,n} q \equiv q(T_{\ell+1,n}) - q(T_{\ell,n}).
\end{aligned}$$

Hence, if we take $\dot{\Pi}_x(\boldsymbol{\xi}) = (\partial_{\Pi_x^M \boldsymbol{\xi}})_x \Pi^M$ and $\ddot{\Pi}_x(\boldsymbol{\xi}, \boldsymbol{\eta}) = (\partial_{\Pi_x^M \boldsymbol{\eta}})_x \dot{\Pi}(\boldsymbol{\xi})$, then, by Taylor's Theorem, we can write

$$p_n(T_{m,n}) - p(T_{m,n}, x, \mathbf{w}_n) = \sum_{\ell=0}^{m-1} \left(A_{\ell,n} - B_{\ell,n} + \tfrac{1}{2} C_{\ell,n} - \tfrac{1}{2} D_{\ell,n} + E_{\ell,n} + F_{\ell,n} \right),$$

where

$$\begin{aligned}
A_{\ell,n} &= \left(\Pi_{p_n(T_{\ell,n})} - \Pi_{p(T_{\ell,n}, x, \mathbf{w}_n)} \right) \Delta_{\ell,n} d\mathbf{w} \\
B_{\ell,n} &= 2^n \left(\int_{I_{\ell,n}} \Pi_{p_n(t)}\, dt \right) \Delta_{\ell,n} q \\
C_{\ell,n} &= \left(\dot{\Pi}_{p_n(T_{\ell,n})}(\Delta_{\ell,n}\mathbf{w}) - \dot{\Pi}_{p(T_{\ell,n}, x, \mathbf{w}_n)}(\Delta_{\ell,n}\mathbf{w}) \right) \Delta_{\ell,n}\mathbf{w} \\
D_{\ell,n} &= \dot{\Pi}_{p_n(T_{\ell,n})}(\Delta_{\ell,n} q) \Delta_{\ell,n}\mathbf{w} \\
E_{\ell,n} &= 8^n \int_{I_{\ell,n}} \left(\int_{T_{\ell,n}}^t (\tau - T_{\ell,n}) \ddot{\Pi}_{p_n(\tau)}(\Delta_{\ell,n} p, \Delta_{\ell,n} p) \Delta_{\ell,n} p\, d\tau \right) dt \\
F_{\ell,n} &= 8^n \int_{I_{\ell,n}} \left(\int_{T_{\ell,n}}^t (\tau - T_{\ell,n}) \ddot{\Pi}_{p(\tau,\mathbf{w}_n)}(\Delta_{\ell,n}\mathbf{w}, \Delta_{\ell,n}\mathbf{w}) \Delta_{\ell,n}\mathbf{w}\, d\tau \right) dt.
\end{aligned}$$

Now set

$$U_{m,n} = \mathbb{E}^{\mathbb{R}_x} \left[\max_{0 \le \ell \le m} \left| p_n(T_{\ell,n}) - p(T_{\ell,n}, x, \mathbf{w}_n) \right|_{\mathbb{R}^N}^2 \right].$$

By reasoning in exactly the same way as we did in § 2.3, one can use Doob's Inequality (cf. (2.24)) to check that there is a constant $C < \infty$ such that

$$\mathbb{E}^{\mathbb{R}_x} \left[\max_{1 \le m \le M} \left| \sum_{\ell=0}^{m-1} A_{\ell,n} \right|_{\mathbb{R}^N}^2 \right] \le 2^{-n} C \sum_{\ell=0}^{M-1} U_{\ell,n}.$$

Further, applications of Schwarz's inequality show that

$$\mathbb{E}^{\mathbb{R}_x}\left[\max_{1\le m\le M}\left|\sum_{\ell=0}^{m-1} C_{\ell,n}\right|_{\mathbb{R}^N}^2\right] \le C2^{-2n}M\sum_{\ell=0}^{M-1} U_{\ell,n},$$

$$\mathbb{E}^{\mathbb{R}_x}\left[\max_{1\le m\le M}\left|\sum_{\ell=0}^{m-1} E_{\ell,n}\right|_{\mathbb{R}^N}^2\right] \le C2^{-2n}M,$$

and

$$\mathbb{E}^{\mathbb{R}_x}\left[\max_{1\le m\le M}\left|\sum_{\ell=0}^{m-1} F_{\ell,n}\right|_{\mathbb{R}^N}^2\right] \le C2^{-2n}M$$

for an appropriate choice of $C<\infty$.

To handle the terms involving the $B_{\ell,n}$'s, we first set (cf. (4.41))

$$\tilde{q}(t) = q(t) + \frac{1}{2}\int_0^t \mathbf{N}_{p(\tau)}\,dt,$$

and then write $B_{\ell,n} = \tilde{B}_{\ell,n} - \bar{B}_{\ell,n}$ where

$$\tilde{B}_{\ell,n} = 2^n \int_{I_{\ell,n}} \Pi_{p_n(t)}\Delta_{\ell,n}\tilde{q}\,dt \;\&\; \bar{B}_{\ell,n} = 2^{n-1}\int_{I_{\ell,n}}\left(\int_0^t \Pi_{p_n(t)}\mathbf{N}_{p(\tau)}\,d\tau\right)dt.$$

Because $\Pi_{p(t)}\mathbf{N}_{p(t)} \equiv \mathbf{0}$, there is a $C<\infty$ such that

$$\left|\bar{B}_{\ell,n}\right|_{\mathbb{R}^N} \le C2^n \iint_{I_{\ell,n}^2} \left|p_n(t)-p(\tau)\right|_{\mathbb{R}^N} d\tau dt.$$

Therefore,

$$\mathbb{E}^{\mathbb{R}_x}\left[\max_{0\le m\le M}\left|\sum_{\ell=0}^{m-1} \bar{B}_{\ell,n}\right|_{\mathbb{R}^N}^2\right] \le C2^{-2n}M\sum_{\ell=0}^{M-1} W_{\ell,n},$$

where

$$W_{\ell,n} \equiv \mathbb{E}^{\mathbb{R}_x}\left[\sup_{\tau,t\in I_{\ell,n}} \left|p_n(t)-p(\tau)\right|_{\mathbb{R}^N}^2\right].$$

Next observe that, by construction, the conditional distribution of $\{\tilde{B}_{\ell,n} : \ell \ge 1\}$ under $\mathbb{R}_x$ given p is that of a sequence of independent, centered Gaussian $\mathbb{R}^N$-valued random variables, the ℓth one of which has covariance

$$H_{\ell,n}(p) \equiv 4^n \iiint_{I_{\ell,n}^3} \Pi_{p_n(t_1)}\Pi_{p(\tau)}^{\perp}\Pi_{p_n(t_2)}\,dt_1\,d\tau\,dt_2.$$

Hence, by Doob's Inequality,

$$\mathbb{E}^{\mathbb{R}_x}\left[\max_{1\le \ell\le M}\left|\sum_{\ell=0}^{m-1}\tilde{B}_{\ell,n}\right|_{\mathbb{R}^N}^2\right]\le 4\sum_{\ell=0}^{M-1}\mathbb{E}^{\mathbb{P}_x^M}\left[\mathrm{Trace} H_{\ell,n}(p)\right].$$

Since

$$\Pi_{p_n(t_1)}\Pi^{\perp}_{p(\tau)}\Pi_{p_n(t_2)}=\left(\Pi_{p_n(t_1)}-\Pi_{p(\tau)}\right)\Pi^{\perp}_{p(\tau)}\left(\Pi_{p_n(t_2)}-\Pi_{p(\tau)}\right),$$

we conclude that

$$\mathbb{E}^{\mathbb{R}_x}\left[\max_{1\le m\le M}\left|\sum_{\ell=0}^{m-1}\tilde{B}_{\ell,n}\right|_{\mathbb{R}^N}^2\right]\le C2^{-n}\sum_{\ell=0}^{M-1}W_{\ell,n}.$$

Summarizing, we have proved so far that

$$\mathbb{E}^{\mathbb{R}_x}\left[\max_{1\le m\le M}\left|\sum_{\ell=0}^{m-1}B_{\ell,n}\right|_{\mathbb{R}^N}^2\right]\le C2^{-n}\left(1+2^{-n}M\right)\sum_{\ell=0}^{M-1}W_{\ell,n}.$$

But, by (2.44) and (4.47),

$$W_{\ell,n}\le C\left(2^{-\frac{n}{2}}+U_{\ell-1,n}\right). \tag{4.48}$$

Thus, we can replace the preceding by

$$\mathbb{E}^{\mathbb{R}_x}\left[\max_{1\le m\le M}\left|\sum_{\ell=0}^{m-1}B_{\ell,n}\right|_{\mathbb{R}^N}^2\right]\le C\left(1+2^{-n}M\right)\left(2^{-\frac{3n}{2}}M+2^{-n}\sum_{\ell=0}^{M-1}U_{\ell,n}\right).$$

The terms $D_{\ell,n}$ are handled in a similar fashion. First we write

$$D_{\ell,n}=\dot{\Pi}_{p_n(T_{\ell,n})}\left(\Delta_{\ell,n}\tilde{q}\right)\Delta_{\ell,n}\mathbf{w}-\frac{1}{2}\left(\int_{I_{\ell,n}}\dot{\Pi}_{p_n(T_{\ell,n})}\left(\mathbf{N}_{p(t)}\right)dt\right)\Delta_{\ell,n}\mathbf{w}.$$

The second term presents no problems: the square of its $L^2(\mathbb{R}_x;\mathbb{R}^N)$-norm is dominated by a constant times 2^{-3n}. Next, by decomposing $\Delta_{\ell,n}\mathbf{w}$ into $\Delta_{\ell,n}p+\Delta_{\ell,n}\tilde{q}-\frac{1}{2}\int_{I_{\ell,n}}\mathbf{N}_{p(t)}\,dt$, we can write

$$\begin{aligned}&\dot{\Pi}_{p_n(T_{\ell,n})}\left(\Delta_{\ell,n}\tilde{q}\right)\Delta_{\ell-1,n}\mathbf{w}\\&\quad=\dot{\Pi}_{p_n(T_{\ell,n})}\left(\Delta_{\ell,n}\tilde{q}\right)\Delta_{\ell,n}p+\dot{\Pi}_{p_n(T_{\ell,n})}\left(\Delta_{\ell,n}\tilde{q}\right)\Delta_{\ell,n}\tilde{q}\\&\qquad-\frac{1}{2}\dot{\Pi}_{p_n(T_{\ell,n})}\left(\Delta_{\ell,n}\tilde{q}\right)\int_{I_{\ell,n}}\mathbf{N}_{p(t)}\,dt.\end{aligned}$$

Again, the last term causes no problem. Moreover,

$$
\begin{aligned}
\mathbb{E}^{\mathbb{R}_x}\left[\left|\dot{\Pi}_{p_n(T_{\ell,n})}(\Delta_{\ell,n}\bar{q})\Delta_{\ell,n}p\right|^2_{\mathbb{R}^N}\right] &\le C\mathbb{E}^{\mathbb{R}_x}\left[\left|\Pi_{p_n(T_{\ell,n})}\Delta_{\ell,n}\bar{q}\right|^2_{\mathbb{R}^N}\left|\Delta_{\ell,n}p\right|^2_{\mathbb{R}^N}\right] \\
&= C\mathbb{E}^{\mathbb{P}^M_x}\left[\left(\operatorname{Trace}\int_{I_{\ell,n}}\Pi_{p_n(T_{\ell,n})}\Pi^{\perp}_{p(t)}\Pi_{p_n(T_{\ell,n})}\,dt\right)\left|\Delta_{\ell,n}p\right|^2_{\mathbb{R}^N}\right] \\
&\le C'2^{-n}\mathbb{E}^{\mathbb{P}^M_x}\left[\sup_{t\in I_{\ell,n}}\left|p(t)-p_n(T_{\ell,n})\right|^2_{\mathbb{R}^N}\left|\Delta_{\ell,n}p\right|^2_{\mathbb{R}^N}\right] \\
&\le 2C'2^{-n}\mathbb{E}^{\mathbb{P}^M_x}\left[\sup_{t\in I_{\ell,n}}\left|p(t)-p(T_{\ell,n})\right|^4_{\mathbb{R}^N}\right] \\
&\qquad + 2C'2^{-n}\mathbb{E}^{\mathbb{P}^M_x}\left[\left|p(T_{\ell,n})-p_n(T_{\ell,n})\right|^2_{\mathbb{R}^N}\left|\Delta_{\ell,n}p\right|^2_{\mathbb{R}^N}\right] \\
&\le C''2^{-3n} + C''2^{-2n}\mathbb{E}^{\mathbb{P}^M_x}\left[\left|p(T_{\ell,n})-p_n(T_{\ell,n})\right|^2_{\mathbb{R}^N}\right] \\
&\le C'''2^{-2n}\left(2^{-\frac{n}{2}} + U_{\ell,n}\right),
\end{aligned}
$$

where we have made repeated application of the reasoning used earlier and, in the last line, have applied (4.48). Finally, in estimating the remaining term, it is useful to note that the $\mathbb{R}_x$-distribution of

$$
(p,q)\in\mathcal{P}(M)\times\mathcal{P}(\mathbb{R}^N)\longmapsto\left(\Pi_{p_n(T_{\ell,n})}\Delta_{\ell,n}\bar{q},\Delta_{\ell,n}\bar{q}\right)\in\mathbb{R}^N\times\mathbb{R}^N
$$

is the same as that of

$$
(p,\boldsymbol{\xi})\in\mathcal{P}(M)\times\mathbb{R}^N\longmapsto\left(\Pi_{p_n(T_{\ell,n})}\circ\Sigma_p\boldsymbol{\xi},\Sigma_p\boldsymbol{\xi}\right)\in\mathbb{R}^N\times\mathbb{R}^N,
$$

under $\mathbb{P}^M_x\times\Gamma$, where Σ_p is the symmetric, non-negative definite square root of $\int_{I_{\ell,n}}\Pi^{\perp}_{p(t)}\,dt$ and Γ is the standard, normal distribution on $\mathbb{R}^N$. In particular,

$$
\begin{aligned}
\mathbb{E}^{\mathbb{R}_x}\left[\left|\dot{\Pi}_{p_n(T_{\ell,n})}(\Delta_{\ell,n}\bar{q})\Delta_{\ell,n}\bar{q}\right|^2_{\mathbb{R}^N}\right] &\le C2^{-n}\mathbb{E}^{\mathbb{P}^M_x}\left[\left\|\Pi_{p_n(T_{\ell,n})}\circ\Sigma_p\right\|^2_{\mathrm{op}}\right]\mathbb{E}^{\Gamma}\left[|\boldsymbol{\xi}|^4\right] \\
&\le C3N^2 2^{-n}\mathbb{E}^{\mathbb{P}^M_x}\left[\operatorname{Trace}\left(\Sigma_p\Pi^2_{p_n(T_{\ell,n})}\Sigma_p\right)\right] \\
&= C3N^2\mathbb{E}^{\mathbb{P}^M_x}\left[\operatorname{Trace}\left(\Pi_{p_n(T_{\ell,n})}\Sigma_p^2\Pi_{p_n(T_{\ell,n})}\right)\right] \\
&\le C'2^{-2n}\mathbb{E}^{\mathbb{P}^M_x}\left[\sup_{t\in I_{\ell,n}}\left|p(t)-p_n(T_{\ell,n})\right|^2\right] \le C''2^{-2n}\left(2^{-\frac{n}{2}}+U_{\ell-1,n}\right),
\end{aligned}
$$

where, once again, we have applied (4.48) at the end.

When we put all these together, we now find that there exists a $C<\infty$ such that

$$
U_{M,n}\le C\left(1+2^{-n}M\right)\left(2^{-\frac{3n}{2}}M+\sum_{\ell=0}^{M-1}U_{\ell,n}\right),
$$

and so, by Lemma 2.23,

$$U_{M,n} \le C2^{-\frac{n}{2}}\left(1+2^{-n}M\right)^2 \exp\Big(C\big(1+2^{-n}M\big)\Big).$$

Finally,

$$\begin{aligned}\sup_{t\in[0,T]} &\big|p_n(t) - p(t,x,\mathbf{w}_n)\big|_{\mathbb{R}^N} \\ &\le 2 \max_{1\le m<2^nT} \big|\Delta_{m,n}p\big|_{\mathbb{R}^N} \vee \big|\Delta_{m,n}\mathbf{w}\big|_{\mathbb{R}^N} \\ &\qquad + \max_{1\le m<2^nT}\Big|p_n\big(T_{m+1,n}\big) - p\big(T_{m+1,n},x,\mathbf{w}_n\big)\Big|_{\mathbb{R}^N}.\end{aligned}$$

In particular, by (1.6) and (2.46) with $\alpha = \frac{1}{4}$,

$$\mathbb{E}^{\mathbb{R}_x}\left[\sup_{t\in[0,T]}\big|p_n(t) - p(t,x,\mathbf{w}_n)\big|^2_{\mathbb{R}^N}\right] \le C(T)2^{-\frac{n}{2}} + 3U_{[2^nT],n},$$

which, together with (4.47) and the estimates just obtained, leads immediately to

$$E^{\mathbb{P}^M_x}\left[\sup_{t\in[0,T]}\big|p_n(t) - p(t)\big|^2_{\mathbb{R}^N}\right] \le C(T)2^{-\frac{n}{2}} \tag{4.49}$$

for an appropriate choice of non-decreasing $T\in[0,\infty) \longmapsto C(T)\in(0,\infty)$.

More about Extrinsic Riemannian Geometry

We continue in this chapter with the same setting as that in § 4.2. Thus, M is a smooth, closed, embedded submanifold of $\mathbb{R}^N$, and we endow M with the Riemannian structure which it inherits from $\mathbb{R}^N$. Our goal is to see how much more of this Riemannian structure we can find reflected by the Brownian motion which was constructed in § 4.2.2.

5.1 Parallel Transport

One of the fundamental concepts in Riemannian geometry is that of *parallel transport*. For a submanifold M of $\mathbb{R}^N$, this concept can be described as follows.

Given an absolutely continuously path $p : [a,b] \longrightarrow M$, one says that $X_{p(b)} \in T_{p(b)}M$ *is obtained from* $X_{p(a)} \in T_{p(a)}M$ *by parallel transport along* p *and writes* $X_{p(b)} = \mathcal{T}_p X_{p(a)}$ if there is an absolutely continuous $t \in [a,b] \longmapsto X(t) \in T_{p(t)}M$ such that

$$\Pi^M_{p(t)} \dot{X}(t) = \mathbf{0} \text{ on } [a,b] \text{ and } X(t) = X_{p(t)} \text{ for } t \in \{a,b\}. \tag{5.1}$$

That is, the only change in $t \rightsquigarrow X(t)$ along p occurs in the direction perpendicular to the tangent space of M.

5.1.1. Parallel Transport along Smooth Paths. We will need the following lemmas in our discussion of existence, uniqueness, and basic properties of parallel transport. In the statement, and elsewhere, we use the notation (cf. (4.28))

$$D^M_{\boldsymbol{\xi}} \equiv \partial_{\Pi^M_x \boldsymbol{\xi}} = \sum_{i=1}^N (\boldsymbol{\xi}, \mathbf{e}_i)_{\mathbb{R}^N} D_i^{\mathbf{e},M} \quad \text{for } \boldsymbol{\xi} \in \mathbb{R}^N. \tag{5.2}$$

5.3 LEMMA. *Given* $x \in M$, *define* $\partial \Pi^M_x \in \mathrm{Hom}(T_xM; \mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N))$ *so that for each* $X_x \in T_xM$, $\partial\Pi^M_x(X_x)$ *is the element of* $\mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N)$ *determined by*[1]

$$\partial\Pi^M_x(X_x)\boldsymbol{\xi} = X_x \Pi^M \boldsymbol{\xi}, \quad \boldsymbol{\xi} \in \mathbb{R}^N.$$

[1] For the benefit of my readers who are well-trained in differential geometry, I feel obliged to say that the "correct" expression for the quantity on the right hand side of the next equation is $\nabla^{\mathbb{R}^N}_{X_x} \partial_{\Pi^M \boldsymbol{\xi}}$, where $\nabla^{\mathbb{R}^N}$ is used to denote (standard Euclidean) covariant differentiation in $\mathbb{R}^N$. For everyone else, the meaning is simply that I am thinking of $x \rightsquigarrow \Pi^M_x \boldsymbol{\xi}$ as a smooth, $\mathbb{R}^N$-valued function on M on which I am acting the directional derivative X_x in the *naive sense*: coordinate by coordinate. Also, I suggest that the interested reader keep in mind the case when M is the unit sphere $\mathbf{S}^{N-1}$ in $\mathbb{R}^N$. In that case, $\Pi^M_x \boldsymbol{\xi} = \boldsymbol{\xi} - (\boldsymbol{\xi}, x)_{\mathbb{R}^N} x$.

Then $\partial\Pi_x^M(X_x)$ *is symmetric. Moreover,*

$$\partial\Pi_x^M(X_x) : T_xM \longrightarrow (T_xM)^\perp \quad \text{and} \quad \partial\Pi_x^M(X_x) : (T_xM)^\perp \longrightarrow T_xM.$$

Equivalently,

$$\Pi^M \circ \partial\Pi^M = \partial\Pi^M \circ (\Pi^M)^\perp \quad \text{and} \quad (\Pi^M)^\perp \circ \partial\Pi^M = \partial\Pi^M \circ \Pi^M.$$

Finally, if $x \in M \longmapsto \mathcal{A}_x^M \in \mathrm{Hom}(\mathbb{R}^N; \mathrm{Hom}(\mathbb{R}^N; \mathbb{R}^N))$ *is given by*

$$\mathcal{A}_x^M(\boldsymbol{\xi}) = \partial\Pi_x^M\big((D_{\boldsymbol{\xi}}^M)_x\big) \circ \big(\Pi_x^M - (\Pi_x^M)^\perp\big),$$

then $\mathcal{A}_x^M(\boldsymbol{\xi})$ *is skew symmetric for all* $x \in M$ *and* $\boldsymbol{\xi} \in \mathbb{R}^N$.

PROOF: The symmetry of $\partial\Pi_x^M(X_x)$ is inherited from that of Π_x^M. To prove the asserted mapping properties of $\partial\Pi_x^M(X_x)$, note that, because Π^M is idempotent,

$$\partial\Pi_x^M(X_x)\boldsymbol{\xi} = X_x\Pi^M\boldsymbol{\xi} = X_x(\Pi^M)^2\boldsymbol{\xi} = \partial\Pi_x^M(X_x)\Pi_x^M\boldsymbol{\xi} + \Pi_x^M\partial\Pi_x^M(X_x)\boldsymbol{\xi},$$

which means $(\Pi_x^M)^\perp \circ \partial\Pi_x^M(X_x)\boldsymbol{\xi} = \partial\Pi_x^M(X_x) \circ \Pi_x^M\boldsymbol{\xi}$. By taking transposes throughout, one also gets $\partial\Pi_x^M(X_x) \circ (\Pi_x^M)^\perp = \Pi_x^M \circ \partial\Pi_x^M(X_x)$.

Given the preceding, the asymmetry of $\mathcal{A}_x^M(\boldsymbol{\xi})$ is easy. Indeed,

$$\begin{aligned}\mathcal{A}_x^M(\boldsymbol{\xi})^\top &= \Pi_x^M\partial\Pi_x^M(D_{\boldsymbol{\xi}}^M) - (\Pi_x^M)^\perp\partial\Pi_x^M(D_{\boldsymbol{\xi}}^M) \\ &= \partial\Pi_x^M\big((D_{\boldsymbol{\xi}}^M)_x\big)(\Pi_x^M)^\perp - \partial\Pi_x^M\big((D_{\boldsymbol{\xi}}^M)_x\big)\Pi_x^M = -\mathcal{A}_x(\boldsymbol{\xi}). \quad \square\end{aligned}$$

5.4 THEOREM. *Given an absolutely continuous* $p : [a,b] \longrightarrow M$, *determine* $t \in [a,b] \longmapsto O_p(t) \in \mathrm{Hom}(\mathbb{R}^N; \mathbb{R}^N)$ *by*

$$\dot{O}_p(t) = \mathcal{A}_{p(t)}^M(\dot{p}(t))O_p(t) \quad \text{with } O_p(a) = I. \tag{5.5}$$

Then $t \rightsquigarrow O_p(t)$ *takes values in the orthogonal group* $O(\mathbb{R}^N)$. *Furthermore, for each* $t \in [a,b]$,

$$\Pi_{p(t)}^M \circ O_p(t) = O_p(t) \circ \Pi_{p(a)}^M \quad \text{and} \quad (\Pi_{p(t)}^M)^\perp \circ O_p(t) = O_p(t) \circ (\Pi_{p(a)}^M)^\perp. \tag{5.6}$$

Finally, if $X_{p(a)} \in T_{p(a)}M$, *then, for each* $t \in [a,b]$, $O_p(t)X_{p(a)}$ *is the one and only element of* $T_{p(t)}M$ *obtained from* $X_{p(a)}$ *by parallel transport along* $p \restriction [a,t]$. *That is*

$$\mathcal{T}_p = O_p(b) \restriction T_{p(a)}M.$$

PROOF: Since $\mathcal{A}_x^M(\boldsymbol{\xi})$ is skew symmetric, it is elementary to check that $O_p(t) \in O(\mathbb{R}^N)$ for each $t \in [a,b]$. Next, to prove the first equality in (5.6), set $B_p(t) = \Pi_{p(t)}^M \circ O_p(t)$, and observe that, by the relations derived in Lemma 5.3,

$$\dot{B}_p(t) = \partial\Pi_{p(t)}^M(\dot{p}(t))O_p(t) - \partial\Pi_{p(t)}^M(\dot{p}(t))(\Pi_{p(t)}^M)^\perp O_p(t) = \mathcal{A}_{p(t)}^M(\dot{p}(t))B_p(t).$$

Since this is exactly the same as the equation satisfied by $t \rightsquigarrow O_p(t) \circ \Pi^M_{p(a)}$ and the two agree when $t = a$, uniqueness for solutions of linear ordinary differential equations guarantees that they agree for all $t \in [a, b]$. Given the first part of (5.6), the second part is immediate.

Finally, we want to check that $X_{p(b)} = O_p(b)X_{p(a)}$ is the one and only element of $T_{p(b)}M$ which is obtained from $X_{p(a)} \in T_{p(a)}M$ by parallel transport along p. To this end, first set $X(t) = O_p(t)X_{p(a)}$, and observe that $t \in [a, b] \longmapsto X(t) \in T_{p(t)}M$ is an absolutely continuous map. In addition, by Lemma 5.3,

$$\dot{X}(t) = \partial\Pi^M_{p(t)}(\dot{p}(t))\Pi^M_{p(t)}X(t) = (\Pi^M_{p(t)})^{\perp}\partial\Pi^M_{p(t)}(\dot{p}(t))X(t),$$

and so $t \rightsquigarrow X(t)$ satisfies (5.1). That is, $X_{p(b)}$ is obtained from $X_{p(a)}$ by parallel transport along p. Hence, it remains to show that there is at most one such $X_{p(b)}$; and for this purpose it suffices to show that $\mathbf{0}$ is the only element of $T_{p(b)}$ which is obtained from $\mathbf{0}$ by parallel transport along p. But if $t \in [a, b] \longmapsto X(t) \in T_{p(t)}M$ is an absolutely continuous map which satisfies the first part of (5.1), then

$$\frac{d}{dt}\big(X(t), X(t)\big)_{\mathbb{R}^N} = 2\big(\dot{X}(t), X(t)\big)_{\mathbb{R}^N} = 0,$$

because $X(t) \in T_{p(t)}M$ and $\Pi^M_{p(t)}\dot{X}(t) = \mathbf{0}$. Hence, $X(a) = \mathbf{0} \implies X(b) = \mathbf{0}$. □

5.1.2. Parallel Transport along Brownian Paths. Our next step is to extend the notion of parallel transport so that it applies to Brownian paths on M, and, in view of Theorem 5.4, this is tantamount to finding out how to define $O_p \in C\big([0, \infty); O(\mathbb{R}^N)\big)$ for "typical" Brownian paths p. To carry this out in a reasonably elegant way, it is best to exploit the fact that $O(\mathbb{R}^N)$ is a Lie group[2] whose Lie algebra $o(\mathbb{R}^N)$ is the $\frac{1}{2}N(N-1)$-dimensional vector space of skew symmetric transformations a on $\mathbb{R}^N$. Thus, there is a linear map taking $a \in o(\mathbb{R}^N)$ to the *right invariant* vector field $\rho(a)$ on $O(\mathbb{R}^N)$ which is given at O by

$$\rho(a)_O f = \frac{d}{d\xi} f\left(e^{\xi a}O\right)\Big|_{\xi=0}.$$

In particular, if $p : [0, \infty) \longrightarrow M$ is absolutely continuous, then (5.5) can be written as

$$\dot{O}_p(t) = \rho\Big(\mathcal{A}^M_{p(t)}(\dot{p}(t))\Big)_{O_p(t)} \quad \text{with } O_p(0) = I.$$

In order to make this situation look like one to which the results in § 4.1 might apply, we introduce the inner product

$$(A, B)_{o(\mathbb{R}^N)} \equiv \operatorname{Trace}(AB^{\top}) = -\operatorname{Trace}(AB)$$

---

[2] The reader who is uncomfortable with Lie groups should just interpret what follows by thinking about $O(\mathbb{R}^N)$ as a submanifold in $\operatorname{Hom}(\mathbb{R}^N; \mathbb{R}^N) \simeq \mathbb{R}^{N^2}$.

on $o(\mathbb{R}^N)$. We can then define the adjoint map $a \in o(\mathbb{R}^N) \longmapsto \Xi_x^M(a) \in \mathbb{R}^N$ by the duality relation

$$\left(\Xi_x^M(a), \boldsymbol{\xi}\right)_{\mathbb{R}^N} = \left(\mathcal{A}_x^M(\boldsymbol{\xi}), a\right)_{o(\mathbb{R}^N)}.$$

Notice that, because $\mathcal{A}_x^M(\boldsymbol{\xi}) = \mathbf{0}$ whenever $\boldsymbol{\xi} \perp T_xM$, Ξ_x^M takes its values in T_xM. Finally, choose $\left\{A_\alpha : 1 \leq \alpha \leq \frac{1}{2}N(N-1)\right\}$ to be a basis in $o(\mathbb{R}^N)$ which is orthonormal with respect to the $o(\mathbb{R}^N)$-inner product, and define the vector fields $(\mathfrak{D}_1^{\mathfrak{c},M}, \dots, \mathfrak{D}_N^{\mathfrak{c},M})$ on $M \times O(\mathbb{R}^N)$ so that

$$(5.7)\quad (\mathfrak{D}_i^{\mathfrak{c},M})_{(x,O)} f = (D_i^{\mathfrak{c},M})_x f(\,\cdot\,, O) + \sum_\alpha \left(\Xi_x^M(A_\alpha), \mathbf{e}_i\right)_{\mathbb{R}^N} \rho(A_\alpha)_O f(x, *).$$

Then (cf. § 3.3), if

$$\mathfrak{p}(t, x, \mathbf{w}_n) = \left(p(t, x, \mathbf{w}_n), O(t, x, \mathbf{w}_n)\right) \quad \text{where } O(t, x, \mathbf{w}_n) \equiv O_{p(\,\cdot\,, x, \mathbf{w}_n)}(t),$$

$\mathfrak{p}(\,\cdot\,, x, \mathbf{w}_n)$ is determined by

$$\dot{\mathfrak{p}}(t, x, \mathbf{w}_n) = \sum_{i=1}^N \left(\mathbf{w}_n(t), \mathbf{e}_i\right)_{\mathbb{R}^N} (\mathfrak{D}_i^{\mathfrak{c},M})_{\mathfrak{p}(t,x,\mathbf{w}_n)} \quad \text{with } \mathfrak{p}(0, x, \mathbf{w}_n) = (x, I).$$

5.8 THEOREM. *Assume that the martingale problem for $\frac{1}{2}\Delta_M$ starting at $x \in M$ admits a solution $\mathbb{P}_x^M$. Then the martingale problem for the operator $\frac{1}{2}\sum_{i=1}^N (\mathfrak{D}_i^{\mathfrak{c},M})^2$ has a unique solution $\widetilde{\mathbb{P}_x^M} \in \mathcal{P}\big(M \times O(\mathbb{R}^N)\big)$ starting at (x, I). In fact, there is a measurable map $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^N) \longmapsto \mathfrak{p}(\,\cdot\,, x, \mathbf{w}) \in \mathcal{P}\big(M \times O(\mathbb{R}^N)\big)$ whose $\mu_{\mathbb{R}^N}$-distribution is $\widetilde{\mathbb{P}_x^M}$, and*

$$\mathfrak{p}(\,\cdot\,, x, \mathbf{w}_n) \longrightarrow \mathfrak{p}(\,\cdot\,, x, \mathbf{w}) \text{ in } \mathcal{P}\big(M \times O(\mathbb{R}^N)\big) \text{ for } \mu_{\mathbb{R}^N}\text{-almost every } \mathbf{w}.$$

In particular, $\mathbf{w} \rightsquigarrow p(\,\cdot\,, x, \mathbf{w})$ under $\mu_{\mathbb{R}^N}$ is a Brownian motion on M starting at x, and, for $\mu_{\mathbb{R}^N}$-almost every $\mathbf{w}$,

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \|O_{p(\,\cdot\,,x,\mathbf{w})} - O_{p(\,\cdot\,,x,\mathbf{w}_n)}(t)\|_{\mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N)} = 0, \quad T \in [0, \infty).$$

Hence, $\mathfrak{p} \in C\big([0,\infty); M \times O(\mathbb{R}^N)\big) \longmapsto p \in C\big([0,\infty); M\big)$ under $\widetilde{\mathbb{P}_x^M}$ is a Brownian motion on M starting at x and, for $\widetilde{\mathbb{P}_x^M}$-almost every $\mathfrak{p} \in \mathcal{P}\big(M \times O(\mathbb{R}^N)\big)$,

$$\Pi_{p(t)}^M \circ O(t) = O(t) \circ \Pi_x^M \text{ and } (\Pi_{p(t)}^M)^\perp \circ O(t) = O(t) \circ (\Pi_x^M)^\perp, \quad t \in [0, \infty).$$

PROOF: The only question is whether explosion can occur with positive $\mu_{\mathbb{R}^N}$-probability. However, because the $\mathfrak{D}_i^{\mathfrak{c},M}$'s and their derivatives are bounded on $B_{\mathbb{R}^N}(\mathbf{0}, r) \times O(\mathbb{R}^N)$ for every $r > 0$, one sees that explosion for $\mathfrak{p}(\,\cdot\,, x, \mathbf{w})$

is $\mu_{\mathbb{W}^N}$-equivalent to explosion for $p(\,\cdot\,,x,\mathbf{w})$. Thus, existence of a solution to the martingale problem for $\frac{1}{2}\Delta_M$ starting at x implies the existence of $\widetilde{\mathbb{P}_x^M}$ as well. □

5.1.3. An Internal Description. [3] Although it certainly should be true that parallel transport along a Brownian path is a measurable function of the Brownian path, the preceding description leaves room for doubt. In order to remove any doubt, we will use the results in § 4.3. Namely, we will show that (cf. (4.44) and (5.5))

$$\lim_{n\to\infty}\sup_{t\in[0,T]}\big\|O(t)-O_{p_n}(t)\big\|_{\mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N)}=0\quad \widetilde{\mathbb{P}_x^M}\text{-almost surely}\tag{5.9}$$

for each $(T,x)\in(0,\infty)\times M$.

Given § 4.3, the proof of (5.9) is quite easy. In the first place, just as we did there, we may and will assume that M is compact. In particular, (4.49) holds. Next, we again use the probability measure (cf. (4.46)) $\mathbb{R}_x$ on $\mathcal{P}(M)\times\mathcal{P}(\mathbb{R}^N)$ which is the joint distribution of the Brownian motion p and the normal process q. In view of the last part of Theorem 5.8, it is then clear that (5.9) is equivalent to checking that

$$\lim_{n\to\infty}\sup_{t\in[0,T]}\big\|O_{p(\,\cdot\,,x,\mathbf{w}_n)}(t)-O_{p_n}(t)\big\|_{\mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N)}=0\quad \mathbb{R}_x^M\text{-almost surely}$$

when $\mathbf{w}(\,\cdot\,)\equiv p(\,\cdot\,)-p(0)+q(\,\cdot\,)$ and $p(\,\cdot\,,x,\mathbf{w}_n)$ is defined as in (4.36). But (with the notation in § 4.3.1)

$$\begin{aligned}
&O_{p(\,\cdot\,,x,\mathbf{w}_n)}(t)-O_{p_n}(t)\\
&\quad=2^n\sum_{m=1}^{\infty}\int_{T_{m-1,n}\wedge t}^{T_{m,n}\wedge t}\Big(\mathcal{A}^M_{p(\tau,x,\mathbf{w}_n)}\big(\Pi^M_{p(\tau,x,\mathbf{w}_n)}\Delta_{m,n}\mathbf{w}\big)O_{p(\,\cdot\,,x,\mathbf{w}_n)}(\tau)\\
&\qquad\qquad\qquad-\mathcal{A}^M_{p_n}\big(\Pi^M_{p_n(\tau)}\Delta_{m,n}\mathbf{w}\big)O_{p_n}(\tau)\Big)\,d\tau\\
&\qquad+2^n\sum_{m=1}^{\infty}\int_{T_{m-1,n}\wedge t}^{T_{m,n}\wedge t}\mathcal{A}^M_{p_n(\tau)}\big(\Pi^M_{p_n(\tau)}\Delta_{m,n}q\big)O_{p_n}(\tau)\,d\tau.
\end{aligned}$$

Hence, if

$$U_n(T,x)\equiv\mathbb{E}^{\mathbb{R}_x}\left[\sup_{t\in[0,T]}\big\|O_{p(\,\cdot\,,x,\mathbf{w}_n)}(t)-O_{p_n}(t)\big\|^2_{\mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N)}\right],$$

then, by precisely the same reasoning as was used in § 4.3.1, one can first show that

$$U_n(T,x)\le C(T)\left(2^{-\frac{n}{2}}+2^{-n}\sum_{1\le m<2^nT}U_n(T_{m,n},x)\right),$$

[3] This section covers material which will not play an important role in what follows. It is included for pedants who, like the author, might have been curious about such matters.

and then apply Gronwall's inequality to get

$$\mathbb{E}^{R_x}\left[\sup_{t\in[0,T]}\left\|O_{p(\cdot,x,\mathbf{w}_n)}(t)-O_{p_n}(t)\right\|^2_{\mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N)}\right]\le C(T)2^{-\frac{n}{2}} \tag{5.10}$$

for appropriate choices of non-decreasing $T\in[0,\infty)\longmapsto C(T)\in[0,\infty)$.

5.2 Riemannian Connection, Covariant Derivatives, & Curvature

In the modern literature on Riemannian geometry, it is customary to discuss parallel transport after introducing a Riemannian connection. However, in the setting here, when M is a submanifold of $\mathbb{R}^N$ and we are giving it the Riemannian structure which it inherits from $\mathbb{R}^N$, it seemed preferable to begin with the more intuitive notion of parallel transport and defer the explicit mention of a connection until later. Indeed, now we can define the Riemannian connection on M in terms of parallel transport; and once we have the Riemannian connection described, we will use it to define the Riemannian curvature.

5.2.1. Riemannian Connection and Covariant Derivatives. Given a vector $X_x\in T_xM$, choose a smooth path $p:\mathbb{R}\longrightarrow M$ so that $p(0)=x$ and $\dot p(0)=X_x$, and define the *covariant derivative* $\nabla^M_{X_x}Y$ of the vector field Y on M by the formula

$$\nabla^M_{X_x}Y=\lim_{t\searrow 0}\frac{1}{t}\left(\mathcal{T}^{-1}_{p\restriction[0,t]}Y_{p(t)}-Y_x\right). \tag{5.11}$$

To see that this limit exists and is independent of the choice of p, we use Theorem 5.4 to rewrite the right hand side as[4]

$$\frac{d}{dt}O_p(t)^\top Y_{p(t)}\bigg|_{t=0}=X_x(Y)-\partial\Pi^M_x(X_x)Y_x.$$

That is,

$$\nabla^M_{X_x}Y=X_x(Y)-\partial\Pi^M_x(X_x)Y_x=\Pi^M_xX_x(Y), \tag{5.12}$$

where the last equality comes from the observation that

$$X_x(Y)=X_x\big(\Pi^MY\big)=\Pi^M_xX_x(Y)+\partial\Pi^M_x(X_x)Y_x. \tag{5.13}$$

Notice that, from the second equality in (5.12), it is immediate that, for vector fields X, Y, and Z on M, one has the Leibnitz rule

$$Z(X,Y)_{\mathbb{R}^N}=\big(\nabla^M_ZX,Y\big)_{\mathbb{R}^N}+\big(X,\nabla^M_ZY\big)_{\mathbb{R}^N}. \tag{5.14}$$

[4] Again, the "correct" expression for the first quantity on the right hand side of the following is $\nabla^{\mathbb{R}^N}_{X_x}Y$.

The operator ∇^M is called the *Riemannian connection* which M inherits from $\mathbb{R}^N$. More precisely, because it satisfies (5.14) and is *torsion free* in the sense that, for any pair of vector fields X and Y on M,[5]

$$[X,Y] = \nabla^M_X Y - \nabla^M_Y X, \tag{5.15}$$

∇^M is the *Levi-Civita* connection corresponding to this Riemannian structure.

A look at (5.11) should be sufficient to convince one that we can covariantly differentiate a smooth vector field which is defined along a smooth curve. To be precise, given a smooth curve $p : [a,b) \longrightarrow M$, we say that the map $t \in [a,b) \longmapsto Y(t) \in T_{p(t)}M$ is a *smooth vector field along* p if $t \rightsquigarrow Y(t)$ is smooth when thought of as a map from $[a,b)$ into $\mathbb{R}^N$. Then, given a smooth p and a smooth vector field $t \rightsquigarrow Y(t)$ along p, (5.11) tells us to define the *covariant derivative of* Y *along* p to be the vector field $\frac{DY}{dt}$ along p given by

$$\frac{DY}{dt}(t) = \lim_{\tau \searrow 0} \frac{1}{\tau}\Big(\mathcal{T}^{-1}_{p\restriction[t,t+\tau]}Y(t+\tau) - Y(t)\Big), \quad t \in [a,b). \tag{5.16}$$

An equivalent, but less easily motivated, expression for the same thing is

$$\frac{DY}{dt}(t) = \sum_{k=1}^N \Big(\tfrac{d}{dt}\big(Y(t),\mathbf{e}_k\big)_{\mathbb{R}^N}\,(D_k^{\iota,M})_{p(t)} + \big(Y(t),\mathbf{e}_k\big)_{\mathbb{R}^N}\,\nabla^M_{\dot p(t)} D_k^{M,\iota}\Big).$$

Either from (5.16) directly or from this latter expression together with (5.14), it is an elementary matter to check that for any smooth vector fields $t \rightsquigarrow X(t)$ and $t \rightsquigarrow Y(t)$ along p,

$$\frac{d}{dt}\big(X(t),Y(t)\big)_{\mathbb{R}^N} = \left(\frac{DX}{dt}(t), Y(t)\right)_{\mathbb{R}^N} + \left(X(t), \frac{DY}{dt}(t)\right)_{\mathbb{R}^N}. \tag{5.17}$$

Finally, suppose that $(s,t) \in [a,b] \times [c,d] \longmapsto M$ is a smooth map. Then, because

$$\frac{\partial^2 f \circ p}{\partial s \partial t} = \frac{\partial^2 f \circ p}{\partial t \partial s}, \quad f \in C^2(M;\mathbb{R}),$$

(5.15) leads to

$$\frac{D\dot p}{ds}(s,t) = \frac{Dp'}{dt}(s,t), \quad \text{where } \dot p(s,t) = \frac{\partial p}{\partial t}(s,t) \text{ and } p'(s,t) = \frac{\partial p}{\partial s}(s,t). \tag{5.18}$$

5.2.2. The Second Fundamental Form & Minimal Submanifolds. We devote this subsection to a short digression on the subject of minimal submanifolds. In particular, we will check here the assertion, made in § 4.2.4, that the vector $\mathbf{N}_x$ in (4.41) is the mean curvature normal to M at x.

[5] Recall that the commutator $[X,Y]$ of two vector fields on M is again a vector field on M, and use (5.12).

To understand this assertion, we first need to introduce the second fundamental form; and for this purpose it is best for us to adopt, at least for the moment, the correct notation for coordinate by coordinate differentiation of vector fields on $\mathbb{R}^N$. Namely, as we have mentioned before (at least in footnotes) what we have been denoting by $X(Y)$ is actually the covariant derivative of Y with respect to X relative to the Levi-Civita connection on $\mathbb{R}^N$. Thus, the correct notation for $X(Y)$ is $\nabla_X^{\mathbb{R}^N} Y$. In particular, when we adopt this notation, (5.13) becomes

$$\nabla_{X_x}^{\mathbb{R}^N} Y = \nabla_{X_x}^M Y + \partial\Pi_x^M(X_x)Y_x. \tag{5.19}$$

Equivalently, $\nabla_{X_x}^M Y$ and $\partial\Pi_x^M(X_x)Y_x$ are, the projections of $\nabla_{X_x}^{\mathbb{R}^N} Y$ onto, respectively, T_xM and $(T_xM)^\perp$. But this means (cf. §3 in Chapter VII of [24]) that

$$(X_x, Y_x) \in T_xM \times T_xM \longmapsto \partial\Pi_x^M(X_x)Y_x \in (T_xM)^\perp$$

is the *second fundamental form* on M at x. Notice that the second fundamental form is symmetric in the sense that

$$\partial\Pi_x^M(X_x)Y_x = \partial\Pi_x^M(Y_x)X_x \quad \text{for } X_x,\, Y_x \in T_xM. \tag{5.20}$$

To see this, simply observe that

$$\Pi^M[X,Y] = [X,Y] = X\Pi^M Y - Y\Pi^M X = \partial\Pi^M(X)Y - \partial\Pi^M(Y)X + \Pi^M[X,Y].$$

Finally, the trace of the second fundamental form is what differential geometers (cf. page 34 in [24]) call the mean curvature normal.

To relate these considerations to $\mathbf{N}_x$, recall that $\mathbf{N}_x$ arose in § 4.2.4 as the first order part of Δ_M when we expressed it in terms of Euclidean coordinates in $\mathbb{R}^N$. Thus,

$$\mathbf{N}_x = \sum_{k=1}^N D_k^{\mathfrak{e},M}(D_k^{\mathfrak{e},M}) = \sum_{k=1}^N \partial\Pi_x^M\big((D_k^{\mathfrak{e},M})_x\big)\partial_k^{\mathfrak{e}}. \tag{5.21}$$

Since the right hand side is obviously independent of the choice of $\mathfrak{e}$, we are free to choose $\mathfrak{e}$ so that (4.32) holds, in which case it becomes clear that the right hand side is nothing but the trace of the second fundamental form as described in the preceding paragraph. In other words, we have now verified the assertion that $\mathbf{N}_x$ is the mean curvature normal to M at x.

Having established that $\mathbf{N}_x$ is the mean curvature vector at x, we have also justified the claim in § 4.2.4 about minimal submanifolds. Namely, a *minimal submanifold* M in $\mathbb{R}^N$ is one for which the mean curvature vector vanishes at each point. Equivalently, M is minimal if and only if

$$[\Delta_M f](x) = \sum_{i,j=1}^N \big(\Pi_x \mathbf{e}_i, \mathbf{e}_j\big)_{\mathbb{R}^N} (\partial_i^{\mathfrak{e}})_x \partial_j^{\mathfrak{e}} f.$$

Hence, an alternative characterization is that M is minimal if and only if the restriction to M of every affine function is Δ_M-harmonic. In any case, because $\mathbf{N} \equiv 0$, Theorem 4.43 says that *Brownian motion on a closed, embedded, minimal submanifold M in $\mathbb{R}^N$ never explodes.*

Next, suppose that M is a closed, embedded, minimal submanifold of $\mathbb{R}^N$ and that $f(x) = \varphi(|x|_{\mathbb{R}^N})$ for some smooth $\varphi : (0,\infty) \longrightarrow \mathbb{R}$. Then

$$\begin{gathered}[\Delta_M f](x) = \beta(x)\varphi''\big(|x|_{\mathbb{R}^N}\big) + \frac{\dim(M) - \beta(x)}{|x|_{\mathbb{R}^N}}\varphi'\big(|x|_{\mathbb{R}^N}\big)\\ \text{with } \beta(x) \equiv \frac{|\Pi_x x|^2_{\mathbb{R}^N}}{|x|^2_{\mathbb{R}^N}}.\end{gathered} \tag{5.22}$$

In particular, $\Delta_M\big(|x|^2_{\mathbb{R}^N}\big) = 2\dim(M)$, and so, for any $R \in (0,\infty)$ and $x \in M$, we have that both

$$\Big(p(t\wedge\zeta_R), \mathcal{F}_t, \mathbb{P}^M_x\Big) \quad\text{and}\quad \Big(\big|p(t\wedge\zeta_R)\big|^2_{\mathbb{R}^N} - \dim(M)\big(t\wedge\zeta_R(p)\big), \mathcal{F}_t, \mathbb{P}^M_x\Big)$$

are martingales when $\zeta_R(p) = \inf\{t \ge 0 : |p(t)|_{\mathbb{R}^N} \ge R\}$. Hence, by Doob's inequality,

$$\begin{aligned}\mathbb{E}^{\mathbb{P}^M_x}\left[\sup_{t\in[0,T]}\big|p(t)\big|^2_{\mathbb{R}^N}\right] &= \lim_{R\to\infty}\mathbb{E}^{\mathbb{P}^M_x}\left[\sup_{t\in[0,T]}\big|p(t\wedge\zeta_R)\big|^2_{\mathbb{R}^N}\right]\\ &\le 4\lim_{R\to\infty}\mathbb{E}^{\mathbb{P}^M_x}\left[\big|p(T\wedge\zeta_R)\big|^2_{\mathbb{R}^N}\right] = 4|x|^2_{\mathbb{R}^N} + 4\dim(M)T.\end{aligned}$$

After an application of Lebesgue's Dominated Convergence Theorem, we can now conclude that

$$\mathbb{E}^{\mathbb{P}^M_x}\left[\big|p(t)\big|^2_{\mathbb{R}^N}\right] = |x|^2_{\mathbb{R}^N} + \dim(M)t, \quad (t,x) \in [0,\infty)\times M.$$

This, of course, proves the well-known fact that there are no compact, minimal submanifolds in $\mathbb{R}^N$. Moreover, by refining the same line of reasoning, we arrive at the following interesting fact about the behavior of Brownian motion on closed minimal submanifolds.

5.23 THEOREM. *Assume that M is a closed, embedded, minimal submanifold of $\mathbb{R}^N$ and that $m \equiv \dim(M) \ge 3$. Then, for each $x \in M$,*

$$\lim_{t\to\infty}\big|p(t)\big|_{\mathbb{R}^N} = \infty \quad \mathbb{P}^M_x\text{-almost surely.}$$

PROOF: When $M = \mathbb{R}^N$, this result is classical (cf. Corollary 7.2.14 in [35]). Thus we will assume that $M \ne \mathbb{R}^N$. In fact, we assume that $\mathbf{0} \notin M$. Next take $f(x) = |x|^{2-m}$ and apply (5.22) to see that $\Delta_M|x|^{2-m}_{\mathbb{R}^N} \le 0$. Hence, $|p(t)|^{2-m}_{\mathbb{R}^N}$ is a bounded, non-negative $\mathbb{P}^M_x$-supermartingale, which means, by Doob's Martingale Convergence Theorem (cf. Theorem 7.1.16 in [35]), that

$Y = \lim_{t\to\infty} |p(t)|_{\mathbb{R}^N}^{2-m}$ exists both $\mathbb{P}_x^M$-almost surely and in $L^1(\mathbb{P}_x^M;\mathbb{R})$. Thus, all that remains is to prove that $Y = 0$ $\mathbb{P}_x^M$-almost surely. To this end, we again use (5.22), this time with $f(x) = |x|_{\mathbb{R}^N}$, and thereby arrive first at $\Delta_M |x|_{\mathbb{R}^N} \geq (m-1)|x|_{\mathbb{R}^N}^{-1}$ and then at

$$\mathbb{E}^{\mathbb{P}_x^M}\left[|p(t)|_{\mathbb{R}^N}\right] \geq |x|_{\mathbb{R}^N} + \frac{m-1}{2}\int_0^t \mathbb{E}^{\mathbb{P}_x^M}\left[|p(\tau)|_{\mathbb{R}^N}^{-1}\right].$$

In particular,

$$\mathbb{P}_x^M(Y>0)>0 \implies \varliminf_{t\to\infty} \frac{\mathbb{E}^{\mathbb{P}_x^M}\left[|p(t)|_{\mathbb{R}^N}\right]}{t} > 0,$$

which, because of Schwarz's inequality, contradicts $\mathbb{E}^{\mathbb{P}_x^M}\left[|p(t)|_{\mathbb{R}^N}^2\right] = |x|_{\mathbb{R}^N}^2 + mt$. $\square$

The content of Theorem 5.23 can be summarized by the statement that Brownian motion on a closed, embedded, minimal submanifold of $\mathbb{R}^N$ of dimension at least three is *transient*. In that the same is true about Euclidean Brownian motion, this may not come as a big surprise. On the other hand, one should not be too careless about ones predictions. There are two dimensional, closed, embedded, minimal submanifolds of $\mathbb{R}^N$ on which the Brownian motion is transient.

5.2.3. Riemannian and Ricci Curvature. Given vector fields X, Y, and Z on M, the vector field $R(X,Y)Z$ is defined in terms of the Riemannian connection by the formula[6]

$$R(X,Y)Z = \left[\nabla_X^M, \nabla_Y^M\right]Z - \nabla_{[X,Y]}^M Z. \tag{5.24}$$

Although we have defined it in terms of vector fields, it turns out that the vector $\big(R(X,Y)Z\big)_x$ depends on X_x, Y_x, and Z_x alone, and not on the vector fields X, Y, and Z anywhere else. To make this fact apparent, we use (5.12) to write

$$\begin{aligned} R(X,Y)Z &= \Pi^M\Big(X\big(\Pi^M Y(Z)\big)\Big) - \Pi^M\Big(Y\big(\Pi^M X(Z)\big)\Big) - \Pi^M\big([X,Y](Z)\big) \\ &= \Pi^M \circ \partial\Pi^M(X)Y(Z) - \Pi^M \circ \partial\Pi^M(Y)X(Z). \end{aligned}$$

The preceding (together with the first part of Lemma 5.3) implies that

$$R(X,Y)Z = \partial\Pi^M(X)\circ(\Pi^M)^\perp Y(Z) - \partial\Pi^M(Y)\circ(\Pi^M)^\perp X(Z).$$

[6] There appears to be a good deal of arbitrariness about the sign convention for the Riemannian curvature. Some books (e.g., [2] and [8]), use the opposite sign convention. My own choice was made on democratic principles: of the differential geometry books in my possession, the majority have made the same choice. I have no convictions on the subject and mention it only to comfort readers who may have been exposed to the other convention.

But, by the second equality in (5.12), we know that $(\Pi^M)^\perp X(Z) = \partial\Pi^M(X)Z$ and $(\Pi^M)^\perp Y(Z) = \partial\Pi^M(Y)Z$, and therefore we now arrive at the formula

$$R(X_x, Y_x)Z_x \equiv \big(R(X,Y)Z\big)_x = \big[\partial\Pi^M_x(X_x), \partial\Pi^M_x(Y_x)\big]Z_x. \tag{5.25}$$

In this connection, refer to the discussion at the end of §5.1.1, and suppose that $(s,t) \in [a,b] \times [c,d] \longmapsto Z(s,t) \in T_{p(s,t)}M$ is smooth. Then, by the same reasoning as was used in the derivation of (5.18), (5.24) leads to

$$\frac{D^2 Z}{\partial s \partial t} - \frac{D^2 Z}{\partial t \partial s} = R(p'(s,t), \dot{p}(s,t))Z(s,t) \tag{5.26}$$

The four tensor determined by

$$(X_x, Y_x, Z_x, W_x) \in (T_x M)^4 \longmapsto \big(R(X_x, Y_x)Z_x, W_x\big)_{\mathbb{R}^N} \in \mathbb{R}$$

is called the *Riemann curvature* , and the expression in (5.25) reveals most of its basic symmetries. Specifically,

$$\begin{aligned}
&\text{(R1)} && R(X_x, Y_x) = -R(Y_x, X_x)\\
&\text{(R2)} && R(X_x, Y_x)Z_x + R(Y_x, Z_x)X_x + R(Z_x, X_x)Y_x = 0\\
&\text{(R3)} && \big(R(X_x, Y_x)Z_x, W_x\big)_{\mathbb{R}^N} = -\big(R(X_x, Y_x)W_x, Z_x\big)_{\mathbb{R}^N}\\
&\text{(R4)} && \big(R(X_x, Y_x)Z_x, W_x\big)_{\mathbb{R}^N} = \big(R(Z_x, W_x)X_x, Y_x\big)_{\mathbb{R}^N}
\end{aligned}$$

The first and third of these are immediate from (5.25). To see the second, we need to use (5.20). Namely, write out the left hand side of (R2) using (5.25) and expanding each of the commutators, and use (5.20) to get the required cancellations. Finally, to check (R4),[7] use (5.25) to write the left hand side of (R4) as

$$\big(\partial\Pi^M_x(Y_x)Z_x, \partial\Pi^M_x(X_x)W_x\big)_{\mathbb{R}^N} - \big(\partial\Pi^M_x(X_x)Z_x, \partial\Pi^M_x(Y_x)W_x\big)_{\mathbb{R}^N},$$

and then apply (5.20) again. The relation in (R2) is sometimes called the *first Bianci identity* .

Because we are dealing with Δ_M, most of our applications of these quantities will not involve the full Riemann curvature tensor but only a contraction of it known as the *Ricci curvature*. Namely, given an orthonormal basis $(E_1, \dots, E_{\dim(M)})$ in $T_x M$, we define $\mathrm{Ric}_x \in \mathrm{Hom}(T_x M; T_x M)$ so that

$$\big(\mathrm{Ric}_x X_x, Y_x\big)_{\mathbb{R}^N} = -\sum_{m=1}^{\dim(M)} \big(R(E_m, X_x)E_m, Y_x\big)_{\mathbb{R}^N}. \tag{5.27}$$

[7] It turns out that (R4) is a consequence of (R1), (R2), and (R3). The first step is to take the inner product of (R2) with W_x, next repeat the first step for each cyclic permutation of (X_x, Y_x, Z_x, W_x), and finally add the resulting four equations and make repeated use of (R1) and (R3).

Clearly, this definition does not depend on the choice of orthonormal basis. In fact, by expanding each $(D_i^{\mathfrak{e},M})_x$ in terms of $(E_1, \dots, E_{\dim(M)})$, one sees that an alternative expression for Ric_x is

$$(5.28)\qquad \left(\mathrm{Ric}_x X_x, Y_x\right)_{\mathbb{R}^N} = -\sum_{m=1}^{N} \Big(R\big((D_i^{\mathfrak{e},M})_x, X_x\big)(D_i^{\mathfrak{e},M})_x, Y_x\Big)_{\mathbb{R}^N}$$

for any orthonormal basis $\mathfrak{e}$ in $\mathbb{R}^N$. Finally, notice that, by (R4), Ric_x is a symmetric transformation on T_xM.

5.3 The Distance Function and Explosion

Associated with any Riemannian metric on a manifold is a notion of *Riemannian distance*. In the case of a submanifold M of $\mathbb{R}^N$ with the Riemannian metric which it inherits from $\mathbb{R}^N$, the Riemannian distance between x, $y \in M$ is given by

$$\mathrm{dist}^M(x,y) \equiv \inf\left\{\int_0^1 |\dot q(t)|_{\mathbb{R}^N}\, dt : q \in C^1\big([0,1]; M\big) \right.$$
$$\left. \text{with } q(0) = x \;\&\; q(1) = y\right\}.$$

As is easily verified, dist^M is a metric on M. Moreover, when (as we assume throughout) M is closed and embedded in $\mathbb{R}^N$, the Hopf–Rinow Theorem (cf. Theorem 4.1 in [24]) guarantees that this metric is complete. Furthermore, because $\int_0^1 |\dot q(t)|_{\mathbb{R}^N}\, dt$ does not change when q is re-parameterized, we need only consider q's for which $t \rightsquigarrow |\dot q(t)|_{\mathbb{R}^N}$ is constant, which, together with elementary compactness considerations familiar in variational calculus (cf. [30]), leads to the conclusion that, for each pair of points x, y from the same connected component of M, there exists at least one $q_0 \in C^1\big([0,1]; M\big)$ such that $q_0(0) = x$, $q_0(1) = y$, and

$$(5.29)\qquad \begin{aligned} \mathrm{dist}^M(x,y)^2 &= \int_0^1 |\dot q_0(t)|^2_{\mathbb{R}^N}\, dt \\ &= \inf\left\{\int_0^1 |\dot q(t)|^2_{\mathbb{R}^N}\, dt : q(0) = x \;\&\; q(1) = y\right\}, \end{aligned}$$

In fact, from the Euler equation associated with the variational problem, it is not hard to deduce that q_0 must be smooth and have *no acceleration* in the sense that $\frac{D\dot q_0}{dt}(t) = 0$ for all $t \in (0,1)$. We will call such a minimizing curve q_0 a *minimal geodesic*. In particular, by (5.17), the *speed* $|\dot q_0(t)|_{\mathbb{R}^N}$ of a minimal geodesic is constantly equal to $\mathrm{dist}^M\big(q(0), q(1)\big)$.

In this section we will show how the distance function can be used to develop a non-explosion criterion for the Brownian motion on M.

5.3.1. Derivatives of the Distance Function. Assume that M is connected, and choose and fix a reference point $x_0 \in M$. Clearly the function $x \in M \longmapsto \operatorname{dist}^M(x, x_0) \in [0,\infty)$ is continuous. On the other hand, even when $M = \mathbb{R}^N$, it fails to have well-defined first derivatives at x_0. To get a feeling for what is going on more generally, notice that

$$\text{(5.30)} \qquad \operatorname{dist}(\,\cdot\,, x_0) \text{ differentiable at } x \implies \left|\operatorname{grad}^M_x \operatorname{dist}^M(\,\cdot\,, x_0)\right|_{\mathbb{R}^N} = 1.$$

To see this, first observe that, for any smooth curve $p : [0,\infty) \longrightarrow M$, the triangle inequality says that

$$\left|\operatorname{dist}^M\big(p(t), x_0\big) - \operatorname{dist}^M\big(p(0), x_0\big)\right| \le \operatorname{dist}^M\big(p(t), p(0)\big) \le \int_0^t \left|\dot p(\tau)\right|_{\mathbb{R}^N} d\tau.$$

Hence, if $\operatorname{dist}^M(\,\cdot\,, x_0)$ is differentiable at x and $X_x \in T_xM$, then we get

$$\left|X_x \operatorname{dist}^M\big(\,\cdot\,, x_0)\big)\right| \le |X_x|_{\mathbb{R}^N}$$

by choosing p so that $p(0) = x$ and $\dot p(0) = X_x$. In other words, we now know that the gradient of $\operatorname{dist}^M(\,\cdot\,, x_0)$ cannot have length more than 1. To prove the opposite inequality, first let $x \in M \setminus \{x_0\}$ be given, and choose $q : [0,1] \longrightarrow M$ to be a minimal geodesic from x_0 to x. Because it is minimizing, an easy argument shows that, for each $s \in (0,1)$, $t \in [0,1] \longmapsto q(st)$ is a minimal geodesic from x_0 to $q(s)$. Thus

$$\operatorname{dist}^M(x, x_0) - \operatorname{dist}^M\big(q(s), x_0\big) = \int_s^1 \left|\dot q(\tau)\right|_{\mathbb{R}^N} d\tau = (1-s)\operatorname{dist}^M(x, x_0),$$

and so

$$\big(\operatorname{grad}^M_x \operatorname{dist}^M(\,\cdot\,, x_0), \dot q(1)\big)_{\mathbb{R}^N} = \operatorname{dist}^M(x, x_0) = |\dot q(1)|_{\mathbb{R}^N}$$

when $\operatorname{grad}^M \operatorname{dist}^M(\,\cdot\,, x_0)$ exists at x. Thus (5.30) is now proved when $x \neq x_0$. To complete, we will show that $\operatorname{dist}^M(\,\cdot\,, x_0)$ is non-differentiable at x_0. To this end, let $q : [0,1] \longrightarrow M$ be any minimal geodesic with $q(0) = x_0$ and $|\dot q(0)|_{\mathbb{R}^N} = 1$. Then, by repeating the argument just given, we find that $\operatorname{dist}^M\big(q(s), x_0\big) = s$. Hence, if $\operatorname{dist}^M(\,\cdot\,, x_0)$ were differentiable at x_0, we would have

$$\operatorname{grad}^M_{x_0} \operatorname{dist}^M(\,\cdot\,, x_0) = \dot q(0) \equiv X_0 \neq \mathbf{0}.$$

But this is impossible. Indeed, choose any smooth $p : [0,\infty) \longrightarrow M$ with $p(0) = x_0$ and $\dot p(0) = -X_0$, and derive the contradiction:

$$0 \le \lim_{t \searrow 0} \frac{\operatorname{dist}^M\big(p(t), x_0\big)}{t} = \Big(\operatorname{grad}^M_{x_0} \operatorname{dist}^M(\,\cdot\,, x_0), -X_0\Big) = -|X_0|^2 = -1.$$

Knowing (5.30), one sees that $\operatorname{dist}^M(\,\cdot\,, x_0)$ must be non-differentiable at any x where $\operatorname{dist}^M(\,\cdot\,, x_0)$ takes a locally extreme value. This is precisely what goes wrong at $x = x_0$, where $\operatorname{dist}^M(\cdot, x_0)$ takes its minimum, and the same thing can happen elsewhere. For example, this certainly will be the case when M is compact, since $\operatorname{dist}^M(\cdot, x_0)$ must then achieve a maximum at some $x \neq x_0$. As a consequence, it is sensible, from the outset, to consider generalized (i.e., distributional) derivatives of $\operatorname{dist}^M(\cdot, x_0)$.

Throughout the discussion which follows, $g : M \longrightarrow [0, \infty)$ will be given by

$$g(x) = F \circ U(x) \quad \text{where } U(x) \equiv \operatorname{dist}^M(x, x_0)^2 \tag{5.31}$$

and F is a smooth, non-decreasing, non-negative function on $[0, \infty)$. Let $\mathfrak{e} = (\mathbf{e}_1, \dots, \mathbf{e}_N)$ be an orthonormal basis in $\mathbb{R}^N$. We begin by showing that, for each $1 \le k \le N$,

$$\left| \int_M (D_k^{\mathfrak{e},M})^{*\lambda_M} \varphi\, g\, d\lambda_M \right| \le 2 \int_M |\varphi| \sqrt{U}\, F' \circ U\, d\lambda_M, \quad \varphi \in C_c^\infty(M; \mathbb{R}). \tag{5.32}$$

To this end, define $(s, x) \in \mathbb{R} \times M \longrightarrow p_k(s, x) \in M$ so that $p_k(\cdot, x)$ is the integral curve of $D_k^{\mathfrak{e},M}$ passing through x at $s = 0$. Then, by the results in §4.1.2, especially (4.16),

$$\int_M (D_k^{\mathfrak{e},M})^{*\lambda_M} \varphi\, g\, d\lambda_M = \lim_{s \searrow 0} \frac{1}{s} \int_M \varphi(x) \Big(g\big(p_k(s, x)\big) - g(x) \Big)\, \lambda_M(dx).$$

Since,

$$g(y) - g(x) = \big(U(y) - U(x)\big) \int_0^1 F'\big((1-t)U(x) + tU(y)\big)\, dt$$

and, by the triangle inequality,

$$\begin{aligned} |U(y) - U(x)| &= \big(U(y)^{\frac12} + U(x)^{\frac12}\big)\big|U(y)^{\frac12} - U(x)^{\frac12}\big| \\ &\le \big(U(y)^{\frac12} + U(x)^{\frac12}\big) \operatorname{dist}^M(x, y), \end{aligned}$$

(5.32) follows from $\operatorname{dist}^M\big(p_k(s, x), x\big) \le |s|$.

We next want to estimate $\Delta_M g$, in the sense of distributions, from above. For this purpose, it is useful to note that, because $\Delta_M \restriction C_c^\infty(M; \mathbb{R})$ is symmetric in $L^2(\lambda_M; \mathbb{R})$, (4.33) is equivalent to

$$\Delta_M = \sum_{k=1}^N \big((D_k^{\mathfrak{e},M})^{*\lambda_M}\big)^2.$$

Hence, by another application of (4.16),

$$\begin{aligned} &\int_M \Delta_M \varphi\, g\, d\lambda_M = \lim_{s \searrow 0} \int_M \varphi(x) G(s, x)\, \lambda_M(dx) \\ &\text{where } G(s, x) \equiv \frac{1}{s^2} \sum_{k=1}^N \Big(g\big(p_k(s, x)\big) + g\big(p_k(-s, x)\big) - 2g(x) \Big), \end{aligned} \tag{5.33}$$

and so, what we are seeking is an upper bound on $\lim_{s \searrow 0} G(s, x)$.

5.34 LEMMA. *Given $x \in M$, define*

$$
\begin{aligned}
\kappa(x) = \sup\Big\{ -\big(\mathrm{Ric}_y X_y, X_y\big)_{\mathbb{R}^N} &: \mathrm{dist}^M(y, x_0) \le \mathrm{dist}^M(x, x_0) \\
&\quad \text{and } X_y \in T_yM \text{ with } |X_y|_{\mathbb{R}^N} \le 1 \Big\},
\end{aligned} \tag{5.35}
$$

and set

$$\beta(x) = \sqrt{\frac{\kappa(x)U(x)}{\dim(M)}}.$$

Then, as $s \searrow 0$,

$$G(s,x) \le 4U(x)F'' \circ U(x) + 2\dim(M)\big(\beta(x)\coth\beta(x)\big)F' \circ U(x) + \mathcal{O}(s)$$

where the $\mathcal{O}(s)$ is uniform on compact subsets of M, and we take $\beta\coth\beta \equiv 1$ when $\beta = 0$.

PROOF: Let $x \in M$ be given. Choose $q(\,\cdot\,,x)$ to be a minimal geodesic from x to x_0, set (cf. Lemma (5.1)) $\mathbf{e}_k(t,x) = O_{q(\,\cdot\,,x)}(t)\mathbf{e}_k$ for $t \in [0,1]$, and take $\boldsymbol{\xi}_k(t,x) = \alpha(t,x)\mathbf{e}_k(t,x)$, where $\alpha(t,x) = 1 - t$ if $\kappa(x) = 0$ and[8]

$$\alpha(t,x) = \cosh\big(\beta(x)t\big) - \coth\big(\beta(x)\big)\sinh\big(\beta(x)t\big)$$

otherwise. Now define $(s,t) \in \mathbb{R} \times [0,1] \longmapsto Q_k(s,t,x) \in M$ so that, for each $t \in [0,1]$, $Q_k(\,\cdot\,,t,x)$ is the integral curve of $D^M_{\boldsymbol{\xi}_k(t,x)}$ which passes through $q(t,x)$ at time $s = 0$. By (5.29),

$$
\begin{aligned}
&\frac{g\big(p_k(s,x)\big) + g\big(p_k(-s,x)\big) - 2g(x)}{s^2} \\
&\quad \le \frac{1}{s^2}\Bigg[F\left(\int_0^1 |\dot{Q}_k(s,t,x)|^2_{\mathbb{R}^N}\,dt\right) \\
&\qquad\qquad + F\left(\int_0^1 |\dot{Q}_k(-s,t,x)|^2_{\mathbb{R}^N}\,dt\right) - 2F\left(\int_0^1 |\dot{Q}_k(0,t,x)|^2_{\mathbb{R}^N}\,dt\right)\Bigg],
\end{aligned}
$$

where the $p_k(\,\cdot\,,x)$'s are defined as in the preceding discussion. Hence, after an elementary application of Taylor's Theorem, we conclude that (cf. (5.33)) that

$$
\begin{aligned}
G(s,x) \le F'' \circ U(x) &\sum_{k=1}^N \left(\int_0^1 \frac{d}{ds}|\dot{Q}_k(s,t,x)|^2_{\mathbb{R}^N}\Big|_{s=0}\,dt\right)^2 \\
&+ F' \circ U(x)\sum_{k=1}^N \int_0^1 \frac{d^2}{ds^2}|\dot{Q}_k(s,t,x)|^2_{\mathbb{R}^N}\Big|_{s=0}\,dt + \mathcal{O}(|s|),
\end{aligned}
$$

[8] The reason for the choice of $\alpha(\,\cdot\,,x)$ below will become clear in the final step of the proof.

where the $\mathcal{O}(|s|)$ is uniform as x ranges over compact subsets of M. Thus, everything comes down to showing that

$$\sum_{k=1}^{N}\left(\int_0^1 \frac{d}{ds}\left|\dot{Q}_k(s,t,x)\right|_{\mathbb{R}^N}^2\Big|_{s=0}\,dt\right)^2 = 4U(x) \tag{5.36}$$

while

$$\sum_{k=1}^{N}\int_0^1 \frac{d^2}{ds^2}\left|\dot{Q}_k(s,t,x)\right|_{\mathbb{R}^N}^2\Big|_{s=0}\,dt \le 2\dim(M)\beta(x)\coth\beta(x). \tag{5.37}$$

To prove (5.36), first observe that, by (5.17) and (5.18),

$$\begin{aligned}\frac{d}{ds}\left|\dot{Q}_k(s,t,x)\right|_{\mathbb{R}^N}^2\Big|_{s=0} &= 2\int_0^1\left(\frac{D\dot{Q}_k}{\partial s}(0,t,x),\dot{q}(t,x)\right)_{\mathbb{R}^N}dt\\ &= 2\int_0^1\left(\frac{DQ'_k}{\partial t}(0,t,x),\dot{q}(t,x)\right)_{\mathbb{R}^N}dt.\end{aligned}$$

Next, note that

$$Q'_k(0,t,x) = \left(D^M_{\xi_k(t,x)}\right)_{q(t,x)} = \alpha(t,x)\mathcal{T}_{q(\cdot,x)\restriction[0,t]}\left(D_k^{\mathfrak{c},M}\right)_x, \tag{5.38}$$

since (cf. (5.6) and the last part of Theorem 5.4)

$$\Pi^M_{q(t,x)}\circ O_{q(\cdot,x)}(t)\mathbf{e}_k = O_{q(\cdot,x)}(t)\circ\Pi_x\mathbf{e}_k.$$

At the same time, because $\frac{D\dot{q}}{dt}(t,x)\equiv\mathbf{0}$, $\dot{q}(t,x)=\mathcal{T}_{q(\cdot,x)\restriction[0,t]}\dot{q}(0,x)$. Thus,

$$\left|\dot{q}(0,x)\right|_{\mathbb{R}^N} = \operatorname{dist}^M(x,x_0)\ \ \&\ \left(\frac{DQ'_k}{\partial t}(0,t,x),\dot{q}(t,x)\right) = \dot{\alpha}(t,x)\big(\mathbf{e}_k,\dot{q}(0,x)\big)_{\mathbb{R}^N},$$

from which (5.36) is clear.

The derivation of (5.37) is similar, only this time we have to use (5.26) as well. Namely, use (5.17), (5.18), and (5.26) to justify the computation:

$$\begin{aligned}\frac{d^2}{ds^2}\left|\dot{Q}_k(s,t,x)\right|_{\mathbb{R}^N}^2\Big|_{s=0} &= 2\left|\frac{D\dot{Q}_k}{\partial s}(0,t,x)\right|_{\mathbb{R}^N}^2 + 2\left(\frac{D^2\dot{Q}_k}{\partial s^2}(0,t,x),\dot{q}(t,x)\right)_{\mathbb{R}^N}\\ &= 2\left|\frac{DQ'_k}{\partial t}(0,t,x)\right|_{\mathbb{R}^N}^2 + 2\left(\frac{D^2Q'_k}{\partial t\partial s}(0,t,x),\dot{q}(t,x)\right)_{\mathbb{R}^N}\\ &\quad + 2\Big(R\big(Q'_k(0,t,x),\dot{q}(t,x)\big)Q'_k(0,t,x),\dot{q}(t,x)\Big)_{\mathbb{R}^N}.\end{aligned}$$

Next, because $Q'_k(0,t,x) = \alpha(t,x)\Pi^M_{q(t,x)}\mathbf{e}_k(t,x)$ and $\big(\mathbf{e}_1(t,x),\dots,\mathbf{e}_N(t,x)\big)$ is an orthonormal basis in $\mathbb{R}^N$ while $U(x) = |\dot q(t,x)|^2_{\mathbb{R}^N}$ for each $(t,x) \in [0,1]\times M$, (5.28) says that

$$\sum_{k=1}^N \Big(R\big(Q'_k(0,t,x),\dot q(t,x)\big)Q'_k(0,t,x),\dot q(t,x)\Big)_{\mathbb{R}^N}$$
$$= -\alpha(t,x)^2\Big(\mathrm{Ric}_{q(t,x)}\dot q(t,x),\dot q(t,x)\Big)_{\mathbb{R}^N} \le \alpha(t,x)^2\kappa(x)U(x).$$

At the same time, we know that (cf. (5.38))

$$\sum_{k=1}^N \left|\frac{DQ'_k}{\partial t}(0,t,x)\right|^2_{\mathbb{R}^N} = \dot\alpha(t,x)^2\sum_{k=1}^N \big|\Pi^M_x\mathbf{e}_k\big|^2_{\mathbb{R}^N} = \dim(M)\dot\alpha(t,x)^2.$$

In order to complete the program, we note that, by (5.17) and the fact that $\frac{D\dot q}{dt}(t,x) = \mathbf{0}$,

$$\left(\frac{D^2Q'_k}{\partial t\partial s}(0,t,x),\dot q(t,x)\right)_{\mathbb{R}^N} = \frac{d}{dt}\left(\frac{DQ'_k}{\partial s}(0,t,x),\dot q(t,x)\right)_{\mathbb{R}^N},$$

which, because $\alpha(0,x) = 1$ and $\alpha(1,x) = 0$, means that

$$\int_0^1 \left(\frac{D^2Q'_k}{\partial t\partial s}(0,t,x),\dot q(t,x)\right)_{\mathbb{R}^N} dt$$
$$= -\left(\frac{DQ'_k}{\partial s}(0,0,x),\dot q(0,x)\right)_{\mathbb{R}^N} = -\Big(\nabla^M_{(D_k^{\mathfrak{e},M})_x}D_k^{\mathfrak{e},M},\dot q(0,x)\Big)_{\mathbb{R}^N}.$$

Finally, notice that (cf. (5.21))

$$\sum_{k=1}^N \nabla^M_{(D_k^{\mathfrak{e},M})_x}D_k^{\mathfrak{e},M} = \Pi^M_x\sum_{k=1}^N D_k^{\mathfrak{e},M}(D_k^{\mathfrak{e},M}) = \Pi^M_x\mathbf{N}_x = \mathbf{0}$$

since $\mathbf{N}_x \perp T_xM$.

After combining the preceding computations, we find that the left hand side of (5.37) is dominated by

$$2\dim(M)\int_0^1\Big(\dot\alpha(t,x)^2 + \beta(x)^2\alpha(t,x)^2\Big)\,dt = -2\dim(M)\alpha(0,x)\dot\alpha(0,x),$$

where we have integrated once by parts to obtain the last equality. □

Because

$$\beta\coth\beta = \beta + \frac{2\beta}{e^{2\beta}-1} \le 1+\beta,\quad \beta\in[0,\infty),$$

we can combine (5.33) with the estimate in Lemma 5.34 to conclude that (cf. (5.31) and (5.35))

$$\begin{aligned}&\int_M \tfrac{1}{2}\Delta_M\varphi\, g\, d\lambda_M\\ &\leq \int_M \varphi\left[2UF''\circ U + \dim(M)\left(1+\sqrt{\frac{\kappa U}{\dim(M)}}\right)F'\circ U\right]d\lambda_M\end{aligned} \tag{5.39}$$

for all $\varphi \in C_c^\infty(M;[0,\infty))$.

5.3.2. An Intrinsic Non-Explosion Criterion for Brownian Motion. With the preceding estimates at hand, we are now in a position to give our first evidence that *lower bounds on Ricci curvature guarantee moderate growth of Brownian motion.*[9] The chapter which follows will provide corroboration of this general principle.

5.40 Theorem. *Assume that M is connected, and let $x_0 \in M$ be a fixed reference point. If there exists an $\alpha \in \mathbb{N}$ such that*

$$\begin{aligned}\left(\mathrm{Ric}_x X_x, X_x\right)_{\mathbb{R}^N} \geq -\frac{\alpha^2}{\dim(M)}\left(1+\mathrm{dist}^M(x,x_0)^2\right)|X_x|^2_{\mathbb{R}^N}\\ \text{for all } x\in M \text{ and } X_x \in T_xM,\end{aligned} \tag{5.41}$$

then the martingale problem for $\frac{1}{2}\Delta_M$ on M is well-posed. In fact, if $d = \dim(M)$, then for every $\epsilon \in (0,1)$, $(T,x) \in (0,\infty)\times M$, and $R \in (0,\infty)$:

$$\begin{aligned}&\mathbb{P}_x^M\left(\sup_{t\in[0,T]} \mathrm{dist}^M(p(t),x_0) \geq R\right)\\ &\leq \frac{(1+d+\alpha)!}{(1-\epsilon)^{1+d+\alpha}}\exp\left(\epsilon\frac{\mathrm{dist}^M(x,x_0)^2}{2(1-\epsilon)T}\right)e^{-\frac{\epsilon R^2}{2Te^{\alpha T}}},\end{aligned} \tag{5.42}$$

and

$$\begin{aligned}&\mathbb{E}^{\mathbb{P}_x^M}\left[\exp\left(\frac{\epsilon\,\mathrm{dist}^M(p(T),x_0)^2}{2Te^{\alpha T}}\right)\right]\\ &\leq \frac{(1+d+\alpha)!}{(1-\epsilon)^{1+d+\alpha}}\exp\left(\epsilon\frac{\mathrm{dist}^M(x,x_0)^2}{2(1-\epsilon)T}\right).\end{aligned} \tag{5.43}$$

---

[9] In some sense, this fact should not be too surprising. For instance, it is a familiar and important fact in differential geometry that lower bounds on the Ricci curvature guarantee moderate volume growth, and the relationship between volume growth and dispersion of associated stochastic processes is well-established. In fact, A. Grigor'yan, [19] & [20], and others have shown that it is volume growth more than bounds on the Ricci curvature which determine whether explosion occurs.

PROOF: Take (cf. (5.31)) $g = 1+U$. Then (5.32) and (5.39) together with our hypothesis show that Theorem 4.21 applies and tells us that the martingale problem for $\frac{1}{2}\Delta_M$ on M is well-posed.

Next, set $g_n = U^n$ and $g = g_1$. By (5.32) and (5.39), respectively,

$$\max_{1\le k\le N}\left|\int_M (D_k^{c,M})^{*\lambda_M}\varphi\, g_n\, d\lambda_M\right| \le n\int_M |\varphi|\big(g_{n-1}+g_n\big)\, d\lambda_M$$

for all $\varphi \in C_c^\infty(M;\mathbb{R})$ and, when $L \equiv d+\alpha$,

$$\int_M \big(\tfrac{1}{2}\Delta_M\varphi\big) g_n\, d\lambda_M \le \int_M \varphi\Big(2\big(n^2+nL\big)g_{n-1} + n\alpha g_n\Big)\, d\lambda_M$$

for all $\varphi \in C_c^\infty\big(M;[0,\infty)\big)$. Hence, if, for $R\in(0,\infty)$,

$$\zeta_R(p) = \inf\Big\{t\ge 0:\ \operatorname{dist}^M\big(p(t),x_0\big)\ge R\Big\} \tag{5.44}$$

and

$$\psi_n(t,x) = e^{-n\alpha t}\mathbb{E}^{\mathbb{P}_x^M}\Big[g\big(p(t\wedge\zeta_R)\big)^n\Big],$$

then, by Lemma 3.60,

$$\psi_n(T,x) \le g_n(x) + 2n^2\left(1+\frac{L}{n}\right)\int_0^T e^{-\alpha t}\psi_{n-1}(t,x)\, dt.$$

Working by induction on $n\in\mathbb{N}$, we find that

$$\psi_n(T,x) \le \sum_{m=0}^n \frac{A_n}{A_m}\frac{T^{n-m}g(x)^m}{(n-m)!},$$

where $A_0 = 1$ and

$$A_n = 2^n(n!)^2\prod_{\ell=1}^n\left(1+\frac{L}{\ell}\right) \quad \text{for } n\ge 1.$$

By elementary estimation,

$$\frac{A_n}{A_m} \le \begin{cases} (1+L)2^n(n!)^2 n^L & \text{if } m=0 \text{ and } n\ge 1\\ 2^{n-m}\left(\frac{n!}{m!}\right)^2\frac{n^L}{m^L} & \text{if } 1\le m\le n,\end{cases}$$

and therefore

$$\frac{1}{n!}\frac{\psi_n(T,x)}{(2T)^n} \le (1+L)n^L + \sum_{m=1}^n \frac{n^L\, n!}{m^L(m!)^2(n-m)!}\left(\frac{g(x)}{2T}\right)^m$$

for $n \geq 1$. Hence, if $\epsilon \in (0,1)$, then

$$\mathbb{E}^{\mathbb{P}_x^M}\left[\exp\left(\epsilon \frac{g\big(p(T\wedge\zeta_R)\big)}{2Te^{\alpha T}}\right)\right] = \sum_{n=0}^{\infty} \frac{\epsilon^n}{n!}\frac{\psi_n(T,x)}{(2T)^n}$$
$$\leq 1 + (1+L)\sum_{n=1}^{\infty} n^L\epsilon^n + \sum_{m=1}^{\infty}\frac{1}{m^L(m!)^2}\left(\frac{\epsilon g(x)}{2T}\right)^m \sum_{n=m}^{\infty}\frac{n^L\, n!}{(n-m)!}\epsilon^{n-m}.$$

Since

$$1 + (1+L)\sum_{n=1}^{\infty} n^L\epsilon^n \leq 1 + (1+L)\sum_{n=1}^{\infty}\frac{(n+L-1)!}{(n-1)!}\epsilon^n$$
$$= 1 + (1+L)\epsilon\frac{d^L}{d\epsilon^L}\sum_{n=1}^{\infty}\epsilon^{n+L-1} = 1 + (1+L)!\frac{\epsilon}{(1-\epsilon)^{1+L}} \leq \frac{(1+L)!}{(1-\epsilon)^{1+L}}$$

while

$$\sum_{n=m}^{\infty}\frac{n^L\, n!}{(n-m)!}\epsilon^{n-m} \leq \sum_{n=m}^{\infty}\frac{(n+L)!}{(n-m)!}\epsilon^{n-m}$$
$$= \frac{d^{m+L}}{d\epsilon^{m+L}}\sum_{n=0}^{\infty}\epsilon^{n+L} = \frac{(m+L)!}{(1-\epsilon)^{m+L+1}},$$

the preceding leads to

$$\mathbb{E}^{\mathbb{P}_x^M}\left[\exp\left(\epsilon\frac{g\big(p(T\wedge\zeta_R)\big)}{2Te^{\alpha T}}\right)\right]$$
$$\leq \frac{(1+L)!}{(1-\epsilon)^{1+L}}\left(1 + \sum_{m=1}^{\infty}\frac{1}{m!}\left(\frac{\epsilon g(x)}{(1-\epsilon)2T}\right)^m \frac{(m+L)!}{(1+L)!m^L\, m!}\right)$$
$$\leq \frac{(1+L)!}{(1-\epsilon)^{1+L}}\exp\left(\frac{\epsilon g(x)}{(1-\epsilon)2T}\right).$$

Because we already know that there is no explosion, it is clear that (5.43) follows from this after one lets $R \to \infty$. As for (5.42), simply note that

$$\mathbb{P}_x^M\left(\sup_{t\in[0,T]}\operatorname{dist}^M\big(x_0,p(t)\big) \geq R\right)$$
$$\leq \exp\left(-\frac{\epsilon R^2}{2Te^{\alpha T}}\right)\mathbb{E}^{\mathbb{P}_x^M}\left[\exp\left(\epsilon\frac{g\big(p(T\wedge\zeta_R)\big)}{2Te^{\alpha T}}\right)\right]. \quad \square$$

S.-T. Yau [46] was the first person to prove non-explosion criteria from lower bounds on the Ricci curvature, although he proved his result under the assumption that the Ricci curvature is uniformly bounded from below. On the other hand, even though Yau was the first to use it in the way he did, the estimate on which his argument rests, namely (5.39), was proved earlier by E. Calabi [4] under the condition that $\mathrm{Ric} \geq 0$ everywhere.

5.3.3. A Comparison of Explosion Criteria. It is interesting to compare the criterion in Theorem 5.40 with the one in Theorem 4.43. There are two respects in which the one in Theorem 5.40 improves on the earlier one. First, and perhaps most significant, because the Ricci curvature is an intrinsic quantity (i.e., it does not depend on the particular realization of M), the criterion in Theorem 5.40 is also completely intrinsic. By contrast, the criterion in Theorem 4.43 is explicitly dependent on the way in which M is embedded in $\mathbb{R}^N$. Secondly, because $|x-x_0|_{\mathbb{R}^N} \leq \operatorname{dist}^M(x,x_0)$, the growth condition in Theorem 5.40 appears to be more lenient than that in Theorem 4.43. To examine the relationship further, note that (cf. (4.41)) Theorem 4.43 says that

$$\sup_{x\in M} \frac{\big(\mathbf{N}_x, x\big)_{\mathbb{R}^N}}{1+|x|^2} < \infty$$

is a sufficient condition for non-explosion. On the other hand, by using (5.25) and (5.20) and recalling the expression for $\mathbf{N}_x$ given in (5.21), one can write

$$\tag{5.45} \begin{gathered} -\big(\mathrm{Ric}_x X_x, X_x\big)_{\mathbb{R}^N} = \big(S_x X_x, X_x\big)_{\mathbb{R}^N} - \big(\partial\Pi_x(X_x)X_x, \mathbf{N}_x\big)_{\mathbb{R}^N} \\ \text{where } S_x = \sum_{k=1}^{N} \partial\Pi_x\big(D_k^{\mathfrak{c},M}\big)^2. \end{gathered}$$

Hence, one can interpret the criterion in Theorem 5.40 as saying that

$$\sup\left\{ \frac{\big(S_x X_x, X_x\big)_{\mathbb{R}^N} - \big(\partial\Pi_x(X_x)X_x, \mathbf{N}_x\big)_{\mathbb{R}^N}}{1+\operatorname{dist}^M(x,x_0)^2} \right. \\ \left. : x \in M \ \&\ X_x \in T_xM \text{ with } |X_x|_{\mathbb{R}^N} \leq 1 \right\} < \infty.$$

In particular, both criteria will be met as soon as one has an estimate which says that, for each $\boldsymbol{\xi} \in \mathbb{R}^N$, $\|\partial\Pi_x(D_{\boldsymbol{\xi}}^M)\|_{\mathrm{op}}$ grows at most linearly as $|x|_{\mathbb{R}^N} \to \infty$. Nonetheless, when one writes the Yau-type criterion in this way, it becomes clear that there are circumstances in which the rather naive criterion in Theorem 4.43 has advantages over the one in Theorem 5.40. This point is made particularly clear when M is minimal, since then $\mathbf{N} \equiv \mathbf{0}$ and therefore $-\mathrm{Ric} = S$. In fact, as we are about to show, minimal submanifolds provide examples for which (5.41) fails dramatically and yet Brownian motion never explodes.

A ready source[10] of embedded, minimal, closed submanifolds is provided by graphs of holomorphic functions. That is, let $F : \mathbb{C} \longrightarrow \mathbb{C}$ be an entire function, and set

$$M = \big\{\big(x_1, x_2, U(x_1,x_2), V(x_1,x_2)\big) : (x_1,x_2) \in \mathbb{R}^2\big\},$$

[10] The suggestion that an example might be found among minimal submanifolds of this sort was made to me by Mike Wolf and Bob Hardt.

where U and V are the real and imaginary parts of F. Clearly, no matter what F is chosen, the corresponding M is a closed, embedded, connected, 2-dimensional submanifold of $\mathbb{R}^4$. Less obvious is the fact that M must be minimal. To see this, let $\Phi : M \longrightarrow \mathbb{R}^2$ be projection onto the first two coordinates: $\Phi(x_1, x_2, x_3, x_4) = (x_1, x_2)$. Then (M, Φ) is a global coordinate system for M. Moreover, if $f(z) = F'(z)$ and u and v denote the real and imaginary parts of f, then an application of the Cauchy–Riemann equations leads to

$$\begin{aligned}(X_1)_x &\equiv \left(\partial_1^\Phi\right)_x = (\partial_1)_{x_1} + u(x_1,x_2)(\partial_3)_{x_3} + v(x_1,x_2)(\partial_4)_{x_4}\\ (X_2)_x &\equiv \left(\partial_2^\Phi\right)_x = (\partial_2)_{x_2} - v(x_1,x_2)(\partial_3)_{x_3} + u(x_1,x_2)(\partial_4)_{x_4}.\end{aligned}$$

In particular, if $\rho(x_1,x_2) = 1 + u(x_1,x_2)^2 + v(x_1,x_2)^2$ and $E_i = \rho^{-\frac{1}{2}} X_i$, then $\big((E_1)_x, (E_2)_x\big)$ is an orthonormal basis in T_xM at each $x \in M$. Hence (cf. § 5.2.2, especially (5.19)),

$$\mathbf{N}_x = \rho^{-1}(\Pi_x^M)^\perp\big((X_1)_x(X_1) + (X_2)_x(X_2)\big).$$

But, because (u, v) also satisfy the Cauchy–Riemann equations, it is easy to check that $X_1(X_1) + X_2(X_2)$ vanishes identically on M, and so we have now verified that M is minimal.

We next want to compute Ric_x. To this end, first note that, because X_1 commutes with X_2,

$$\rho \mathrm{Ric} X_i = \sum_{j=1}^{2} [\nabla^M_{X_i}, \nabla^M_{X_j}] X_j = [\nabla^M_{X_i}, \nabla^M_{X_{i'}}] X_{i'}$$

where i' equals 2 or 1 depending on whether i equals 1 or 2. Next, set $\ell = \log \rho$, $\ell_{,i} = \partial_i \ell$, and, using

$$\nabla_{X_i} X_j = \rho^{-1}\Big(\big(X_i(X_j), X_1\big)_{\mathbb{R}^N} X_1 + \big(X_i(X_j), X_2\big)_{\mathbb{R}^N} X_2\Big),$$

note that

$$\begin{aligned}\nabla^M_{X_1} X_1 &= \frac{1}{2}(\ell_{,1}X_1 - \ell_{,2}X_2)\\ \nabla^M_{X_2} X_1 = \nabla^M_{X_1} X_2 &= \frac{1}{2}(\ell_{,2}X_1 + \ell_{,1}X_2)\\ \nabla^M_{X_2} X_2 &= \frac{1}{2}(-\ell_{,1}X_1 + \ell_{,2}X_2).\end{aligned}$$

Thus, after combining these with the preceding, we arrive at

$$\mathrm{Ric} = -\frac{1}{2\rho}(\Delta_{\mathbb{R}^2}\ell) I.$$

Finally, the right hand side of the preceding can be simplified by making the observation that, because $\rho(x_1, x_2) = 1 + |f(z)|^2$ when $z = x_1 + \sqrt{-1}\,x_2$, an equivalent expression is

$$\mathrm{Ric}_x = -2\frac{|f'(z)|^2}{(1+|f(z)|^2)^3} I_{T_xM} \quad \text{when } z = x_1 + \sqrt{-1}\,x_2.$$

Given the above considerations, it is now clear how to proceed. For example, take $f(z) = \sin e^z$ for $z \in \mathbb{C}$. Because f is entire, there exists a unique entire F such that $F(0) = 0$ and $F' = f$. Let M denote the closed, embedded, minimal submanifold corresponding to this F, and consider the path $t \in [0,\infty) \longmapsto p(t) = \big(t, 0, U(t,0), V(t,0)\big) \in M$ on this manifold. Then

$$\mathrm{dist}^M\big(\mathbf{0}, p(t)\big) \le \int_0^t \sqrt{1 + \left(\sin e^\tau\right)^2}\, d\tau \le 2t,$$

whereas

$$-\mathrm{Ric}_{p(t)} = 2\frac{\left(e^t \cos e^t\right)^2}{\left(1 + (\sin e^t)^2\right)^3} \ge \frac{e^{2t}}{4}\left(\cos e^t\right)^2.$$

By taking $t = \log(n\pi)$, it becomes clear that (5.41) does not come even close to holding for this M. On the other hand, because it is closed, embedded, and minimal, Brownian motion on M never explodes.

5.3.4. Growth Estimate when Ricci Curvature is Bounded Below. In this subsection we will be assuming that there is a $\gamma \in \mathbb{R}$ such that

$$\text{(5.46)} \qquad \big(\mathrm{Ric}_x X_x, X_x\big)_{\mathbb{R}^N} \ge 2\gamma |X_x|^2 \quad \text{for all } x \in M \text{ and } X_x \in T_xM.$$

In particular, by Theorem 5.40, the martingale problem for $\frac{1}{2}\Delta_M$ is well-posed on M. Moreover, because (5.46) frees us from the dependence, imposed by (5.41), on the reference point x_0, we know that estimates like those in Theorem 5.40 must be freed of their dependence on the reference point x_0. In fact, the purpose of the present subsection is to see that estimates like (5.42) and (5.43) can be considerably sharpened when (5.46) holds.

To see what we have in mind, let $x_0 \in M$, $\alpha \in (0,\infty)$, and $\beta \in (0,\infty)$ be given, and consider (cf. (5.31)) the function

$$g = \exp\Big(\alpha\big(\beta^2 + U\big)^{\frac{1}{2}}\Big).$$

By (5.32) and (5.39), we know that, in the sense of distributions, $\left|\mathrm{grad}^M g\right|_{\mathbb{R}^N}$ is bounded on compacts and

$$\tfrac{1}{2}\Delta_M g \le \left(\frac{\alpha^2}{2} + \alpha \dim(M)\big((2\beta)^{-1} + \sqrt{\gamma^-}\,\big)\right) g, \quad \text{where } \gamma^- = (-\gamma) \vee 0.$$

Hence, by Lemma 3.60, for each $R > 0$ (cf. (5.44)),

$$\mathbb{E}^{\mathbb{P}^M_{x_0}}\Big[g\big(p(T\wedge\zeta_R)\big)\Big] \le \exp\left[\alpha\left(\beta + \frac{\dim(M)T}{2\beta}\right) + \alpha\dim(M)\sqrt{\gamma^-}\,T + \frac{\alpha^2 T}{2}\right].$$

In particular, by taking $\beta = \sqrt{\frac{\dim(M)T}{2}}$, we find that

$$\mathbb{E}^{\mathbb{P}^M_{x_0}}\left[\exp\left(\alpha\sqrt{\frac{\dim(M)T}{2} + U\big(p(T\wedge\zeta_R)\big)}\right)\right] \\ \le \exp\left[\alpha\Big(\sqrt{2\dim(M)T} + \dim(M)\sqrt{\gamma^-}\,T\Big) + \frac{\alpha^2 T}{2}\right].$$

Finally, since, for $\lambda \ge 0$ and $\beta > 0$,

$$\int_{\mathbb{R}} e^{\lambda\alpha}e^{-\beta\alpha^2}\,d\alpha \le 2\int_0^\infty e^{\lambda\alpha}e^{-\beta\alpha^2}\,d\alpha \le 2\int_{\mathbb{R}} e^{\lambda\alpha}e^{-\beta\alpha^2}\,d\alpha,$$

we can multiply both sides of the preceding by $\sqrt{\frac{T}{2\pi\epsilon}}\,e^{-\frac{T\alpha^2}{2\epsilon}}$ with $\epsilon \in (0,1)$ and use integration to obtain

$$\mathbb{E}^{\mathbb{P}^M_{x_0}}\left[\exp\left(\frac{\epsilon}{2T}\left[\frac{\dim(M)T}{2} + U\big(p(T\wedge\zeta_R)\big)\right]\right)\right] \\ \le 2(1-\epsilon)^{-\frac{1}{2}}\exp\left[\frac{\epsilon}{2(1-\epsilon)T}\Big(\sqrt{2\dim(M)T} + \dim(M)\sqrt{\gamma^-}\,T\Big)^2\right],$$

which, after elementary manipulation, yields

$$\mathbb{E}^{\mathbb{P}^M_{x_0}}\left[\exp\left(\epsilon\frac{U\big(p(T\wedge\zeta_R)\big)}{2T}\right)\right] \\ \le 2(1-\epsilon)^{-\frac{1}{2}}\exp\left[\epsilon\frac{2\dim(M) + \gamma^-\dim(M)^2 T}{1-\epsilon}\right].$$

As a consequence, we also find that

$$\mathbb{P}^M_{x_0}\left(\sup_{t\in[0,T]}\operatorname{dist}^M\big(x_0, p(t)\big) \ge R\right) \le e^{-\frac{\epsilon R^2}{2T}}\mathbb{E}^{\mathbb{P}^M_{x_0}}\left[\exp\left(\epsilon\frac{U\big(p(T\wedge\zeta_R)\big)}{2T}\right)\right] \\ \le 2(1-\epsilon)^{-\frac{1}{2}}\exp\left[\epsilon\left(-\frac{R^2}{2T} + \frac{2\dim(M) + \gamma^-\dim(M)^2 T}{1-\epsilon}\right)\right].$$

Because there is nothing to distinguish x_0 from any other point in M, we have now proved that (5.46) implies that, for all $x \in M$ and $\epsilon \in (0,1)$,

$$\mathbb{P}^M_x\left(\sup_{t\in[0,T]}\operatorname{dist}^M\big(x, p(t)\big) \ge R\right) \\ \le 2(1-\epsilon)^{-\frac{1}{2}}\exp\left[\epsilon\left(-\frac{R^2}{2T} + \frac{2\dim(M) + \gamma^-\dim(M)^2 T}{1-\epsilon}\right)\right], \tag{5.47}$$

for all $R > 0$, and, after letting $R \nearrow \infty$ and applying Fatou's Lemma,

$$
\begin{aligned}
&\mathbb{E}^{\mathbb{P}_x^M}\left[\exp\left(\epsilon\frac{\operatorname{dist}^M\big(x, p(T)\big)^2}{2T}\right)\right] \\
&\qquad\qquad \leq 2(1-\epsilon)^{-\frac{1}{2}} \exp\left[\epsilon\frac{2\dim(M) + \gamma^{-}\dim(M)^2 T}{1-\epsilon}\right].
\end{aligned}
\tag{5.48}
$$

CHAPTER 6

Bochner's Identity

In § 5.3 we saw that the Ricci curvature has a great deal to say about the behavior of Brownian motion. However, all the evidence given there was based on the appearance (cf. (5.39) and (5.41)) of the lower bounds on the Ricci curvature in upper bounds for the Laplacian of the distance function. In this chapter we will give much more compelling examples of the role which the Ricci curvature plays in computations involving Brownian motion.

Throughout we will be assuming that M is a closed, connected, embedded submanifold of $\mathbb{R}^N$ and that the martingale problem for $\frac{1}{2}\Delta_M$ is well-posed on M.

6.1 The Jacobian Process & Bochner's Identity

Define $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^N) \longmapsto p(\,\cdot\,, *, \mathbf{w}) \in C^{0,\infty}([0,\infty) \times M; M)$ as in Theorem 4.37. Next, given $x \in M$, define the *Jacobian process*

$$\mathbf{w} \in \mathfrak{W}(\mathbb{R}^N) \longmapsto J(\,\cdot\,, *, \mathbf{w}) \in C^{0,\infty}\big([0,\infty) \times M; \mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N)\big)$$

so that (cf. (4.3))

$$J(t,x,\mathbf{w})\boldsymbol{\xi} = p(t,\,*\,,\mathbf{w})_* \Pi_x^M \boldsymbol{\xi}. \tag{6.1}$$

That is, $J(t,x,\mathbf{w})\boldsymbol{\xi}$ is the element of $\mathbb{R}^N$ such that (cf. (5.2))

$$\begin{aligned}(D^M_{\boldsymbol{\xi}})_x f \circ p(t,\,*\,,\mathbf{w}) &= \big(D^M_{J(t,x,\mathbf{w})\boldsymbol{\xi}}\big)_{p(t,x,\mathbf{w})} f \\ &= \big(J(t,x,\mathbf{w})\boldsymbol{\xi}, \mathrm{grad}^M_{p(t,x,\mathbf{w})} f\big)_{\mathbb{R}^N}.\end{aligned}$$

Obviously, for each $(t,x,\mathbf{w}) \in [0,\infty) \times M \times \mathfrak{W}(\mathbb{R}^N)$, $J(t,x,\mathbf{w})$ is a linear map of $\mathbb{R}^N$ into $T_{p(t,x,\mathbf{w})}M$. Our goal in this section is to get a handle on the "average size" of this map.

6.1.1. The Martingale Characterization of the Jacobian Process. The first step in our program is to remark that the size of $J(t,x,\mathbf{w})$ is the same as the size of (cf. Theorem 5.8)

$$\tilde{J}(t,x,\mathbf{w}) = O(t,x,\mathbf{w})^\top \circ J(t,x,\mathbf{w}), \tag{6.2}$$

which, $\mu_{\mathbb{R}^N}$-almost surely, maps $\mathbb{R}^N$ into T_xM for every $t \in [0,\infty)$. In that it maps T_xM into itself, experience provides good reason to supposing that consideration of $\tilde{J}(t,x,\mathbf{w})$ is preferable to that of $J(t,x,\mathbf{w})$, and confirmation of this suspicion will be provided by an interesting expression for the conditional distribution of $\mathbf{w}\rightsquigarrow\tilde{J}(\,\cdot\,,x,\mathbf{w})$ given $\sigma(\{p(t,x,\mathbf{w}) : t \in [0,\infty)\})$. For this purpose, let $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^N)$ be given, determine $p(\,\cdot\,,*\,,\mathbf{w}_n)$ by (cf. (3.17))

$$\dot{p}(t,x,\mathbf{w}_n) = \Pi^M_{p(t,x,\mathbf{w}_n)}\dot{\mathbf{w}}_n \quad \text{with } p(0,x,\mathbf{w}_n) = x,$$

and $J(\,\cdot\,,*\,,\mathbf{w}_n) \in C^{0,\infty}([0,\infty)\times M; \mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N))$ by

$$\dot{J}(t,x,\mathbf{w}_n)\boldsymbol{\xi} = \partial\Pi^M_{p(t,x,\mathbf{w}_n)}\Big(\Pi^M_{p(t,x,\mathbf{w}_n)}J(t,x,\mathbf{w}_n)\boldsymbol{\xi}\Big)\dot{\mathbf{w}}_n(t)$$
$$\text{with } J(0,x,\mathbf{w}_n) = \Pi^M_x\boldsymbol{\xi}.$$

It is then an easy matter to verify that (6.1) holds when $\mathbf{w}$ is replaced by $\mathbf{w}_n$. In particular, $J(t,x,\mathbf{w}_n)$ maps $\mathbb{R}^N$ into $T_{p(t,x,\mathbf{w}_n)}M$; and so, if (cf. Theorem 5.4)

$$\tilde{J}(t,x,\mathbf{w}_n) = O_{p(\,\cdot\,,x,\mathbf{w}_n)}(t)^\top \circ J(t,x,\mathbf{w}_n),$$

then an application of Theorem 5.4 and (5.20) shows that

$$\begin{aligned}\frac{d}{dt}\tilde{J}(t,x,\mathbf{w}_n)\boldsymbol{\xi} &= -O_{p(\,\cdot\,,x,\mathbf{w}_n)}(t)^\top\partial\Pi^M_{p(\,\cdot\,,x,\mathbf{w}_n)}(t)\big(\dot{p}(t,x,\mathbf{w}_n)\big)J(t,x,\mathbf{w}_n)\boldsymbol{\xi}\\ &\quad + O_{p(\,\cdot\,,x,\mathbf{w}_n)}(t)^\top\partial\Pi^M_{p(t,x,\mathbf{w})}\big(J(t,x,\mathbf{w}_n)\boldsymbol{\xi}\big)\dot{\mathbf{w}}_n(t)\\ &= O_{p(\,\cdot\,,x,\mathbf{w}_n)}(t)^\top\circ\partial\Pi^M_{p(t,x,\mathbf{w}_n)}\big(J(t,x,\mathbf{w}_n)\boldsymbol{\xi}\big)\big(\Pi^M_{p(t,x,\mathbf{w}_n)}\big)^\perp\dot{\mathbf{w}}_n(t).\end{aligned}$$

Equivalently, $\tilde{J}(\,\cdot\,,x,\mathbf{w}_n)$ is determined by

$$\frac{d}{dt}\tilde{J}(t,x,\mathbf{w}_n) = \sum_{i=1}^N(\dot{\mathbf{w}}_n,\mathbf{e}_i)_{\mathbb{R}^N}(Y_i^{\mathfrak{c},M})_{(p(t,x,\mathbf{w}_n),O_{p(\,\cdot\,,x,\mathbf{w}_n)}(t),\tilde{J}(t,x,\mathbf{w}_n)\boldsymbol{\xi})}$$
$$\text{with } \tilde{J}(t,x,\mathbf{w}_n)\boldsymbol{\xi} = \Pi^M_x\boldsymbol{\xi},$$

where $Y_i^{\mathfrak{c},M}$ is the vector field on $M\times O(\mathbb{R}^N)\times\mathbb{R}^N$ given by (cf. (5.2))

$$\begin{aligned}(Y_i^{\mathfrak{c},M})_{(x,O,y)}f &= \sum_{j=1}^N\Big(O^\top\Pi^M_x\partial\Pi^M_x\big((D^M_{Oy})_x\big)\mathbf{e}_i,\mathbf{e}_j\Big)_{\mathbb{R}^N}(\partial_j^{\mathfrak{c}})_y f(x,O,*)\\ &= \sum_{j=1}^N\Big(O^\top\nabla^M_{(D^M_{Oy})_x}D_i^{\mathfrak{c},M},\mathbf{e}_j\Big)_{\mathbb{R}^N}(\partial_j^{\mathfrak{c}})_y f.\end{aligned}\tag{6.3}$$

The following theorem is an immediate consequence of Theorem 4.2 and the preceding computations. In its statement, we think of $\mathfrak{D}_i^{\mathfrak{c},M}$ in (5.7) as a vector field on $M\times O(\mathbb{R}^N)\times\mathbb{R}^N$ by taking

$$(\mathfrak{D}_i^{\mathfrak{c},M})_{(x,O,y)}f = (\mathfrak{D}_i^{\mathfrak{c},M})_{(x,O)}f(\,\cdot\,,*\,,y).$$

6.4 THEOREM. *Assume that the martingale problem for* $\frac{1}{2}\Delta_M$ *is well-posed on* M. *Then, for each* $x \in M$ *and* $\boldsymbol{\xi} \in \mathbb{R}^N$, *the* $\mu_{\mathbb{R}^N}$*-distribution of (cf. Theorem 5.8 and (6.2))*

$$\mathbf{w} \in \mathfrak{W}(\mathbb{R}^N) \longmapsto \big(p(\,\cdot\,,x,\mathbf{w}), O(\,\cdot\,,x,\mathbf{w}), \bar{J}(t,x,\mathbf{w})\boldsymbol{\xi}\big) \in C\big(M \times O(\mathbb{R}^N) \times \mathbb{R}^N\big)$$

is the unique solution to the martingale problem for $\frac{1}{2}\sum_{i=1}^N \big(\mathfrak{D}_i^{\mathfrak{e},M} + Y_i^{\mathfrak{e},M}\big)^2$ *starting at* $(x, I, \Pi_x^M \boldsymbol{\xi})$.

6.1.2. A Stochastic Version of Bochner's Identity. The next step in this program is to describe the conditional $\mu_{\mathbb{R}^N}$-distribution of $\mathbf{w} \rightsquigarrow J(\,\cdot\,,x,\mathbf{w})$ given the σ-algebra (cf. Theorem 5.8)

$$\tag{6.5} \begin{gathered} \sigma\Big(\big\{\mathsf{p}(\tau,x,\mathbf{w}) : \tau \in [0,\infty)\big\}\Big)\Big] \\ \text{where } \mathsf{p}(\,\cdot\,,x,\mathbf{w}) \equiv \big(p(\,\cdot\,,x,\mathbf{w}), O(\,\cdot\,,x,\mathbf{w})\big). \end{gathered}$$

Our ability to give such an description relies on our ability to apply the results in §3.4.2 to the present situation. In particular, we will need the calculation:

$$\tag{6.6} \begin{gathered} \frac{1}{2}\sum_{i=1}^N \big(\mathfrak{D}_i^{\mathfrak{e},M} + Y_i^{\mathfrak{e},M}\big)^2 = Y_0 + \frac{1}{2}\sum_{i=1}^N \Big((\mathfrak{D}_i^{\mathfrak{e},M})^2 + (Y_i^{\mathfrak{e},M})^2\Big) \\ \text{where } (Y_0)_{(x,O,y)} \equiv \frac{1}{2}\sum_{i=1}^N \big(\mathfrak{D}_i^{\mathfrak{e},M}(Y_i^{\mathfrak{e},M})\big)_{(x,O,y)} \end{gathered}$$

is independent of the choice of orthonormal basis $\mathfrak{e}$.

Since[1]

$$\big(\mathfrak{D}_i^{\mathfrak{e},M} + Y_i^{\mathfrak{e},M}\big)^2 = \big(\mathfrak{D}_i^{\mathfrak{e},M}\big)^2 + \big(Y_i^{\mathfrak{e},M}\big)^2 + 2\mathfrak{D}_i^{\mathfrak{e},M} \otimes Y_i^{\mathfrak{e},M} + \mathfrak{D}_i^{\mathfrak{e},M}(Y_i^{\mathfrak{e},M}),$$

(6.6) comes down to showing that $\sum_{i=1}^N \mathfrak{D}_i^{\mathfrak{e},M} \otimes Y_i^{\mathfrak{e},M} = \mathbf{0}$. To this end, recall (cf. (5.7)) that $\Xi_x(A) \in T_xM$, and conclude that

$$\begin{aligned} &\sum_{i=1}^N \big(\mathfrak{D}_i^{\mathfrak{e},M} \otimes Y_i^{\mathfrak{e},M}\big)_{(x,O,y)} \\ &\quad= \sum_{i,j,j'} \big(\Pi^M \mathbf{e}_j, \mathbf{e}_i\big)_{\mathbb{R}^N} \Big((\Pi_x^M)^\perp \partial \Pi_x^M (\Pi_x^M O y) O\mathbf{e}_{j'}, \mathbf{e}_i\Big)_{\mathbb{R}^N} \big(\partial_j^{\mathfrak{e}}\big)_x \big(\partial_{j'}^{\mathfrak{e}}\big)_y \\ &\qquad+ \sum_{i,j',\alpha} \big(\Xi_x(A_\alpha), \mathbf{e}_i\big)_{\mathbb{R}^N} \Big((\Pi_x^M)^\perp \partial \Pi_x^M \big((D_O^M y)_x\big) O\mathbf{e}_{j'}, \mathbf{e}_i\Big)_{\mathbb{R}^N} \rho(A_\alpha)_O \big(\partial_{j'}^{\mathfrak{e}}\big)_y \\ &\quad= \sum_{j,j'} \Big(\Pi^M \mathbf{e}_j, (\Pi_x^M)^\perp \partial \Pi_x^M \big((D_{Oy}^M)_x\big) O\mathbf{e}_{j'}\Big)_{\mathbb{R}^N} \big(\partial_j^{\mathfrak{e}}\big)_x \big(\partial_{j'}^{\mathfrak{e}}\big)_y \\ &\qquad+ \sum_{j',\alpha} \big(\Xi_x(A_\alpha), (\Pi_x^M)^\perp \partial \Pi_x^M (\Pi_x^M O y) O\mathbf{e}_{j'}\big)_{\mathbb{R}^N} \rho(A_\alpha)_O \big(\partial_{j'}^{\mathfrak{e}}\big)_y = \mathbf{0}. \end{aligned}$$

[1] This use of the tensor product notation "$\otimes$" may be confusing. What we mean is that if $A = \sum a_i(x,y)\partial_{x^i}$ and $\mathcal{B} = \sum b_j(x,y)\partial_{y^j}$, then $A \otimes \mathcal{B} = \sum a_i b_j \partial_{x^i}\partial_{y^j}$.

6.7 THEOREM[2]. *For each* $(x, O, y) \in M \times O(\mathbb{R}^N) \times \mathbb{R}^N$, *define (cf. (6.3))*

$$Y_i\big((x,O)\big)_y = (Y_i^{\mathfrak{e},M})_{(x,O,y)} \quad \text{when } 1 \le i \le N,$$

and (cf. (6.6))

$$\big(Y_0(x,O)\big)_y = (Y_0)_{(x,O,y)}.$$

Then, for each $\boldsymbol{\xi} \in \mathbb{R}^N$ *and* $\mathfrak{p} \in \mathcal{P}(M \times O(\mathbb{R}^N))$, *there is a unique solution* $\mathbb{Q}_{\boldsymbol{\xi},\mathfrak{p}}$ *to the martingale problem for the (time dependent) operator*

$$t \rightsquigarrow \mathcal{L}_{t,\mathfrak{p}} \equiv Y_0\big(\mathfrak{p}(t)\big) + \frac{1}{2}\sum_{i=1}^{N} Y_i\big(\mathfrak{p}(t)\big)^2$$

starting at $\Pi^M_{p(0)}\boldsymbol{\xi}$. *Moreover,* $\mathfrak{p} \rightsquigarrow \mathbb{Q}_{\boldsymbol{\xi},\mathfrak{p}}$ *is Borel measurable. Finally, if* A *and* B *are Borel measurable subsets of* $\mathcal{P}(M \times O(\mathbb{R}^N))$ *and* $\mathcal{P}(\mathbb{R}^N)$ *respectively, then, for each* $x \in M$ *and* $\boldsymbol{\xi} \in \mathbb{R}^N$, *(cf. Theorem 5.8)*

$$\mu_{\mathbb{R}^N}\Big(\mathfrak{p}(\,\cdot\,, x, \mathbf{w}) \in A \text{ and } \bar{J}(\,\cdot\,, x, \mathbf{w})\boldsymbol{\xi} \in B\}\Big) = \int_A \mathbb{Q}_{\boldsymbol{\xi},\mathfrak{p}}(B)\,\widetilde{\mathbb{P}^M_x}(d\mathfrak{p}).$$

PROOF: In view of (6.6) and Theorem 6.4, all the assertions are essentially immediate from the results in § 3.4, especially Theorem 3.29. In fact, the situation here is less delicate than the one there since, for each $\mathfrak{p}$, it is clear from the form of the vector fields Y_i that the martingale problem for $t \rightsquigarrow \mathcal{L}_{t,\mathfrak{p}}$ has exactly one solution. □

Before the preceding can be put to use, we must compute the action of $\mathcal{L}_{t,\mathfrak{p}}$ on a couple of simple functions, and for this purpose it is best to write out the first and second order parts of $\mathcal{L}_{t,\mathfrak{p}}$ explicitly. That is, we want to write

$$\mathcal{L}_{t,\mathfrak{p}} = \frac{1}{2}\sum_{i,j}^{N} a_{\mathfrak{p}}(t,y)^{ij}\partial_i^{\mathfrak{e}}\partial_j^{\mathfrak{e}} + \sum_{i=1}^{N} b_{\mathfrak{p}}^i(t,y)\partial_i^{\mathfrak{e}}.$$

Clearly, if $\mathfrak{p}(t) = \big(p(t), O(t)\big)$, then

$$a_{\mathfrak{p}}(t,y) = \sum_{i=1}^{N}\Big(O(t)^\top \Pi^M_{p(t)}\partial\Pi^M_{p(t)}\big(X_{\mathfrak{p}}(t,y)\big)\mathbf{e}_i\Big)^{\otimes 2} \tag{6.8}$$

where $X_{\mathfrak{p}}(t,y) \equiv (D^M_{O(t)y})_{p(t)}$. Less obvious is the fact that

$$\mathbf{b}_{\mathfrak{p}}(t,y) = -\frac{1}{2}O^\top(t)\mathrm{Ric}_{p(t)}X_{\mathfrak{p}}(t,y). \tag{6.9}$$

[2] The basic conclusion drawn in and from this theorem was derived, by a quite different line of reasoning, in [13].

Equivalently, what we are claiming is that

$$(*)\qquad 2(Y_0)_{(x,O,y)} + \sum_{i=1}^{N} \left(Y_i^{\epsilon,M}(Y_i^{\epsilon,M})\right)_{(x,O,y)} = -O^\top \mathrm{Ric}_x \Pi_x^M Oy.$$

In the verification of this claim, it is useful to first observe that, for any vector field X on M,

$$(6.10)\qquad (\mathfrak{D}_i^{\epsilon,M})_{(x,O)} F_X = O^\top \left(\nabla^M_{D_i^{\epsilon,M}} X\right)_x \quad \text{when } F_X(x,O) \equiv O^\top X_x.$$

To see this, one simply repeats the derivation of (5.12). That is, choose a smooth $p : \mathbb{R} \longrightarrow M$ so that $p(0) = x$ and $\dot{p}(0) = (D_i^{\epsilon,M})_x$, and note that

$$(\mathfrak{D}_i^{\epsilon,M})_{(x,O)} F_X = \frac{d}{dt} O^\top O_p(t)^\top X_{p(t)}\bigg|_{t=0} = O^\top \left(\nabla^M_{D_i^{\epsilon,M}} X\right)_x.$$

Knowing (6.10) and using the second line of (6.3), we find that

$$\begin{aligned}
2(Y_0^{\epsilon,M})_{(x,O,y)} &= \sum_{i=1}^{N} \left(\mathfrak{D}_i^{\epsilon,M}(Y_i^{\epsilon,M})\right)_{(x,O,y)} \\
&= \sum_{i=1}^{N} O^\top \left(\nabla^M_{D_i^{\epsilon,M}} \nabla^M_{D^M_{Oy}} D_i^{\epsilon,M} + \nabla^M_{D^M_{\mathcal{A}(D_i^{\epsilon,M})Oy}} D_i^{\epsilon,M}\right)_x \\
&= \sum_{i=1}^{N} O^\top \left(\nabla^M_{D_i^{\epsilon,M}} \nabla^M_{D^M_{Oy}} D_i^{\epsilon,M} - \nabla^M_{\nabla^M_{D_i^{\epsilon,M}} D^M_{Oy}} D_i^{\epsilon,M}\right)_x,
\end{aligned}$$

where, in the passage to the third line, we have used the relation

$$\Pi^M \mathcal{A}(D_i^{\epsilon,M})Oy = -\Pi^M \partial\Pi^M (D_i^{\epsilon,M})Oy,$$

which means that $D^M_{\mathcal{A}(D_i^{\epsilon,M})Oy} = -\nabla^M_{D_i^{\epsilon,M}} D^M_{Oy}$. At the same time,

$$\begin{aligned}
\left(Y_i^{\epsilon,M}(Y_i^{\epsilon,M})\right)_{(x,O,y)} &= \sum_{j=1}^{N} \left(O^\top \left(\nabla^M_{D^M_{O(Y_i^{\epsilon,M})}} D_i^{\epsilon,M}\right)_x, \mathrm{e}_j\right)_{\mathbb{R}^N} (\partial_j)_y \\
&= \sum_{j=1}^{N} \left(O^\top \left(\nabla^M_{\nabla^M_{D^M_{Oy}} D_i^{\epsilon,M}} D_i^{\epsilon,M}\right)_x, \mathrm{e}_j\right)_{\mathbb{R}^N} (\partial_j)_y.
\end{aligned}$$

Hence, by (5.15), we now have that

$$\begin{aligned}
&2(Y_0)_{(x,O,y)} + \sum_{i=1}^{N} \left(Y_i^{\epsilon,M}(Y_i^{\epsilon,M})\right)_{(x,O,y)} \\
&\qquad = O^\top \sum_{i=1}^{N} \left(\nabla^M_{D_i^{\epsilon,M}} \nabla^M_{D^M_{Oy}} D_i^{\epsilon,M} - \nabla^M_{[D_i^{\epsilon,M}, D^M_{Oy}]} D_i^{\epsilon,M}\right)_x \\
&\qquad = O^\top \sum_{i=1}^{N} \left(\nabla^M_{D^M_{Oy}} \nabla^M_{D_i^{\epsilon,M}} D_i^{\epsilon,M}\right)_x - O^\top \mathrm{Ric}_x (D^M_{Oy})_x.
\end{aligned}$$

But (cf. (4.41) and (5.21))

$$\sum_{i=1}^{N} \left(\nabla^M_{D_i^{\mathfrak{c},M}} D_i^{\mathfrak{c},M}\right)_x = \Pi_x^M \mathbf{N}_x = \mathbf{0}$$

for all $x \in M$, and therefore (*) has been verified.

6.11 LEMMA. *For each $x \in M$, define $C_x \in \mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N)$ by (cf. (4.41))*

$$(6.12) \qquad \left(\boldsymbol{\eta}, C_x\boldsymbol{\xi}\right)_{\mathbb{R}^N} = \left(\Pi_x^M\boldsymbol{\eta}, \partial\Pi_x^M(D_{\boldsymbol{\xi}}^M)\mathbf{N}_x\right) - 2\left(\Pi_x^M\boldsymbol{\eta}, \mathrm{Ric}_x\Pi_x^M\boldsymbol{\xi}\right)_{\mathbb{R}^N}.$$

Then C_x is symmetric for all $x \in M$, and, for each $\mathfrak{p} = \big(p(\,\cdot\,),O(\,\cdot\,)\big) \in \mathcal{P}\big(M \times O(\mathbb{R}^N)\big)$ and $\boldsymbol{\xi} \in \mathbb{R}^N$,

$$y(t) + \frac{1}{2}\int_0^t O(\tau)^\top \mathrm{Ric}_{p(\tau)}\Pi^M_{p(\tau)}O(\tau)y(\tau)\,d\tau$$

and

$$|y(t)|^2_{\mathbb{R}^N} - \int_0^t \big(O(\tau)y(\tau), C_{p(\tau)}O(\tau)y(\tau)\big)_{\mathbb{R}^N}\,d\tau$$

are (cf. Theorem 6.7) $\mathbb{Q}_{\boldsymbol{\xi},\mathfrak{p}}$-martingales. In particular, if for each $x \in M$, λ_x denotes the largest eigenvalue of C_x, then

$$(6.13) \qquad \mathbb{E}^{\mathbb{Q}_{\boldsymbol{\xi},\mathfrak{p}}}\left[|y(t)|^2_{\mathbb{R}^N}\right] \le |\boldsymbol{\xi}|^2_{\mathbb{R}^N}\exp\left(\int_0^t \lambda_{p(\tau)}\,dt\right).$$

Finally, if $t \in [0,\infty) \longmapsto \hat{J}_{\mathfrak{p}}(t) \in \mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N)$ is determined by the equation

$$(6.14) \qquad \frac{d}{dt}\hat{J}_{\mathfrak{p}}(t) = -\tfrac{1}{2}O(t)^\top\mathrm{Ric}_{p(t)}\Pi^M_{p(t)}O(t)\hat{J}_{\mathfrak{p}}(t) \quad \textit{with } \hat{J}_{\mathfrak{p}}(0) = \Pi^M_{p(0)},$$

then

$$(6.15) \qquad \mathbb{E}^{\mathbb{Q}_{\boldsymbol{\xi},\mathfrak{p}}}\left[y(t)\right] = \hat{J}_{\mathfrak{p}}(t)\boldsymbol{\xi}.$$

PROOF: We begin by showing that

$$(*) \qquad \mathcal{L}_{t,\mathfrak{p}}|y|^2_{\mathbb{R}^N} = \big(O(t)y, C_{p(t)}O(t)y\big)_{\mathbb{R}^N}.$$

But

$$\mathcal{L}_{t,\mathfrak{p}}|y|^2_{\mathbb{R}^N} = \mathrm{Trace}\,a_{\mathfrak{p}}(t,y) + 2\big(\mathbf{b}_{\mathfrak{p}}(t,y), y\big)_{\mathbb{R}^N},$$

and, by (6.9), the second term is equal to (cf. (6.8) and (6.9))

$$-\big(O(t)y, \mathrm{Ric}_{p(t)}\Pi^M_{p(t)}O(t)y\big)_{\mathbb{R}^N} = -\big(\Pi^M_{p(t)}O(t)y, \mathrm{Ric}_{p(t)}\Pi^M_{p(t)}O(t)y\big)_{\mathbb{R}^N},$$

where, in getting the first expression we have used (5.20), and, in the last relation, we have used the fact the $\mathrm{Ric}_x \in \mathrm{Hom}(T_xM;T_xM)$ for all $x \in M$. On the other hand, by (6.8), the first term is equal to

$$\begin{aligned}
&\sum_{i,j=1}^{N}\Big(O(t)^\top\Pi^M_{p(t)}\partial\Pi^M_{p(t)}(X_{\mathrm{p}}(t,y))\mathbf{e}_i,\mathbf{e}_j\Big)^2_{\mathbb{R}^N}\\
&=\sum_{j=1}^{N}\Big|\partial\Pi^M_{p(t)}(X_{\mathrm{p}}(t,y))\Pi^M_{p(t)}O(t)\mathbf{e}_j\Big|^2_{\mathbb{R}^N}=\sum_{j=1}^{N}\Big|\partial\Pi^M_{p(t)}\big((D^M_{O(t)\mathbf{e}_j})_{p(t)}\big)X_{\mathrm{p}}(t,y)\Big|^2_{\mathbb{R}^N}\\
&=\sum_{j=1}^{N}\Big(X_{\mathrm{p}}(t,y),\partial\Pi^M_{p(t)}\big((D^M_{O(t)\mathbf{e}_j})_{p(t)}\big)^2X_{\mathrm{p}}(t,y)\Big)^2_{\mathbb{R}^N}\\
&=-\Big(\Pi^M_{p(t)}O(t)y,\mathrm{Ric}_{p(t)}\Pi^M_{p(t)}O(t)y\Big)_{\mathbb{R}^N}\\
&\qquad+\Big(\partial\Pi^M_{p(t)}\big((D^M_{O(t)y})_{p(t)}\big)\Pi^M_{p(t)}O(t)y,\mathbf{N}_{p(t)}\Big)_{\mathbb{R}^N},
\end{aligned}$$

where we have used (5.45) in the last line. Hence, after combining this with the earlier calculation of the second term, we have now verified the equation for $\mathcal{L}_{t,\mathrm{p}}|y|^2_{\mathbb{R}^N}$.

We next note that, by (6.9), $\mathcal{L}_{t,\mathrm{p}}y = -\frac{1}{2}O(t)^\top\mathrm{Ric}_{p(t)}\Pi^M_{p(t)}O(t)y$. Hence, by standard localization procedure, using cut-off functions and Doob's Stopping Time Theorem, we know that, for each $R > 0$,

$$y(t\wedge\zeta_R)+\frac{1}{2}\int_0^{t\wedge\zeta_R}O(\tau)^\top\mathrm{Ric}_{p(\tau)}\Pi^M_{p(\tau)}O(\tau)y(\tau)\,d\tau$$

and

$$|y(t\wedge\zeta_R)|^2_{\mathbb{R}^N}-\int_0^{t\wedge\zeta_R}\big(O(\tau)y(\tau),C_{p(\tau)}O(\tau)y(\tau)\big)_{\mathbb{R}^N}\,d\tau$$

are $\mathbb{Q}_{\boldsymbol{\xi},\mathrm{p}}$-martingales when ζ_R denotes the first time $t \ge 0$ that $|y(t)|_{\mathbb{R}^N} \ge R$. In particular, from the second of these, we see that

$$\mathbb{E}^{\mathbb{Q}_{\boldsymbol{\xi},\mathrm{p}}}\left[|y(t\wedge\zeta_R)|^2_{\mathbb{R}^N}\right]\le\int_0^t\lambda_{p(\tau)}\mathbb{E}^{\mathbb{Q}_{\boldsymbol{\xi},\mathrm{p}}}\left[|y(\tau\wedge\zeta_R)|^2_{\mathbb{R}^N}\right]d\tau;$$

and therefore, by Gronwall's inequality,

$$\mathbb{E}^{\mathbb{Q}_{\boldsymbol{\xi},\mathrm{p}}}\left[|y(t)|^2_{\mathbb{R}^N}\right]\le\sup_{R\ge0}\mathbb{E}^{\mathbb{Q}_{\boldsymbol{\xi},\mathrm{p}}}\left[|y(t\wedge\zeta_R)|^2_{\mathbb{R}^N}\right]\le|\boldsymbol{\xi}|^2\exp\left(\int_0^t\lambda_{p(\tau)}\,d\tau\right).$$

Knowing this estimate, it becomes clear that we now have more than enough integrability of see that

$$y(t)+\frac{1}{2}\int_0^t O(\tau)^\top\mathrm{Ric}_{p(\tau)}\Pi^M_{p(\tau)}O(\tau)y\,d\tau$$

is a square-integrable $\mathbb{Q}_{\boldsymbol{\xi},\mathrm{p}}$-martingale. Hence, by Doob's Inequality, we find that, for each $T \in [0,\infty)$, $\sup_{t\in[0,T]} |y(t)|^2_{\mathbb{R}^N}$ is $\mathbb{Q}_{\boldsymbol{\xi},\mathrm{p}}$-integrable, and, after combining this with the preceding, we see that

$$|y(t)|^2_{\mathbb{R}^N} - \int_0^t \big(O(\tau)y(\tau), C_{p(\tau)}O(\tau)y(\tau)\big)_{\mathbb{R}^N}\,d\tau$$

is also a $\mathbb{Q}_{\boldsymbol{\xi},\mathrm{p}}$-martingale. □

We can now prove the result alluded to in the heading for this section.

6.16 THEOREM. *Assume that (cf. (6.12)) $C_x \le \alpha I_{\mathbb{R}^N}$ for some $\alpha > 0$ and all $x \in M$. Then (cf. (6.14) and Theorem 5.8)*

$$\mathbb{E}^{\widetilde{\mathbb{P}^M_x}}\left[\|\hat{J}_{\mathrm{p}}(t)\|^2_{\mathrm{op}}\right] = \mathbb{E}^{\mu_{\mathbb{R}^N}}\left[\|\bar{J}(t,x,\mathbf{w})\|^2_{\mathrm{op}}\right] \le Ne^{\alpha t}.$$

Moreover, for any $f \in C^1_{\mathrm{b}}(M;\mathbb{R})$, $\boldsymbol{\xi} \in \mathbb{R}^N$, and $x \in M$, (cf. (6.14))

$$\begin{aligned}(D^M_{\boldsymbol{\xi}})_x \mathbf{P}^M_t f &= \mathbb{E}^{\widetilde{\mathbb{P}^M_x}}\left[\Big(O(t)\hat{J}_{\mathrm{p}}(t)\boldsymbol{\xi}, \mathrm{grad}^M_{p(t)} f\Big)_{\mathbb{R}^N}\right] \\ &= \mathbb{E}^{\mu_{\mathbb{R}^N}}\left[\Big(O(t,x,\mathbf{w})\hat{J}_{\mathrm{p}(\,\cdot\,,x,\mathbf{w})}(t)\boldsymbol{\xi}, \mathrm{grad}^M_{p(t,x,\mathbf{w})} f\Big)_{\mathbb{R}^N}\right].\end{aligned} \tag{6.17}$$

PROOF: In view of the (6.13), only the second part requires comment. In fact, since the second equality in (6.17) is immediate from the definition of $\widetilde{\mathbb{P}^M_x}$, all that is needed is a proof that the first expression in (6,17) is equal to the third, and this is easy. Namely, given $x \in M$ and $\boldsymbol{\xi} \in \mathbb{R}^N$, determine $s \rightsquigarrow q(s)$ by $\frac{d}{ds}q(s) = (D^M_{\boldsymbol{\xi}})_{q(s)}$ with $q(0) = x$. Then, for any $f \in C^1_{\mathrm{b}}(M;\mathbb{R})$,

$$\begin{aligned} f\big(p(t,q(s),\mathbf{w})\big) - f\big(p(t,x,\mathbf{w})\big) &= \int_0^s (D^M_{\boldsymbol{\xi}})_{q(\sigma)} f \circ p(t,\,*\,,\mathbf{w})\,d\sigma \\ &= \int_0^s \Big(O\big(t,q(\sigma),\mathbf{w}\big)\bar{J}(t,x,\mathbf{w})\boldsymbol{\xi}, \mathrm{grad}^M_{p(t,x,\mathbf{w})} f\Big)_{\mathbb{R}^N}\,d\sigma. \end{aligned}$$

Hence, by Fubini's Theorem, Theorem 6.7, and (6.15), we know that

$$\big[\mathbf{P}^M_t f\big]\big(q(s)\big) - \big[\mathbf{P}^M_t f\big](x) = \int_0^s \mathbb{E}^{\mu_{\mathbb{R}^N}}\big[F(\sigma,\mathbf{w})\big]\,d\sigma,$$

where

$$F(\sigma,\mathbf{w}) \equiv \Big(O\big(t,q(\sigma),\mathbf{w}\big)\hat{J}_{\mathrm{p}(\,\cdot\,,q(\sigma),\mathbf{w})}(t)\boldsymbol{\xi}, \mathrm{grad}^M_{p(t,q(\sigma),\mathbf{w})} f\Big)_{\mathbb{R}^N}.$$

Finally, because (6.13) allows us to see that $F(\sigma,\mathbf{w}) \longrightarrow F(0,\mathbf{w})$ in $L^1(\mu_{\mathbb{R}^N};\mathbb{R})$ as $\sigma \to 0$, we are done. □

Remark: It is worth emphasizing that C_x is an essentially *extrinsic* quantity. Indeed, the first term on the right hand side of (6.12) involves nothing but the second fundamental form (cf. § 5.2.2). Thus, there is no way of gaining control over C_x without knowing how M sits inside $\mathbb{R}^N$. On the other hand, when M is well situated in $\mathbb{R}^N$, then there is hope of controlling C_x in terms of Ric_x. For example, if M is a minimal submanifold of $\mathbb{R}^N$, then $C_x = -2\mathrm{Ric}_x$.

6.1.3. The Classical Bochner's Identity. Having attached Bochner's name to this section, it is only fair that we include somewhere a statement that Bochner himself might have recognized. In fact, we will now derive the classical Bochner's identity (cf. (6.19) below) as a corollary[3] of (6.17). As the derivation will show, (6.19) comes from (6.17) by differentiation. That is, (6.17) gives the integrated form of the commutation relation made explicit in (6.19).

6.18 THEOREM. *For any orthonormal basis $\mathfrak{e}$ in $\mathbb{R}^N$ and $f \in C^\infty(M;\mathbb{R})$, (cf. (5.12))*

$$\operatorname{grad}^M \circ \Delta_M f = \sum_{i=1}^N \left(\nabla^M_{D_i^{\mathfrak{e},M}}\right)^2 \operatorname{grad}^M f - \operatorname{Ric} \operatorname{grad}^M f. \tag{6.19}$$

PROOF: To begin with, observe that (6.19) is completely local in the sense that its validity at a point x depends only on the manifold M in a neighborhood of x. Hence, without loss in generality, we may and will assume that M is compact. In particular, the hypotheses of Theorem 6.16 are then met, and therefore (6.17) holds for all $t \in [0,\infty)$.

Now let $x \in M$ and $\boldsymbol{\xi} \in \mathbb{R}^N$ be given, determine $s \rightsquigarrow q(s)$ by $q(0) = x$ and $\frac{d}{ds} q(s) = (D^M_{\boldsymbol{\xi}})_{q(s)}$, and set $u(t,y) = [\mathbf{P}^M_t f](y)$ and

$$U(t,y) = \mathbb{E}^{\widetilde{\mathbb{P}^M_y}}\left[\left(O(t)\Pi^M_{p(t)}\boldsymbol{\xi}, \operatorname{grad}^M_{p(t)} f\right)_{\mathbb{R}^N}\right].$$

for $(t,y) \in [0,\infty) \times M$. Then, by (6.17),

$$u(t,q(s)) - u(t,x) = \int_0^s \Big(U\big(t,q(\sigma)\big) + V\big(t,q(\sigma)\big)\Big)\, d\sigma,$$

where

$$V(t,y) \equiv \mathbb{E}^{\widetilde{\mathbb{P}^M_y}}\left[\left(O(t)\big(\hat{J}_{\mathrm{p}}(t) - \Pi^M_y\big)\boldsymbol{\xi}, \operatorname{grad}^M_{p(t)} f\right)_{\mathbb{R}^N}\right].$$

Next, note that, for $s \geq 0$,

$$\lim_{t\searrow 0} \frac{u\big(t,q(s)\big) - f\big(q(s)\big)}{t} = \lim_{t\searrow 0} \frac{1}{2t} \mathbb{E}^{\mathbb{P}^M_{q(s)}}\left[\int_0^t \Delta_M f\big(p(\tau)\big)\, d\tau\right] = \tfrac{1}{2}\Delta_M f\big(q(s)\big)$$

and that (cf. (6.14))

$$\lim_{t\searrow 0} \frac{1}{t}\int_0^s V\big(t,q(\sigma)\big)\, d\sigma = -\frac{1}{2}\int_0^s \left(\operatorname{Ric}_{q(\sigma)}\Pi^M_{q(\sigma)}\boldsymbol{\xi}, \operatorname{grad}^M_{q(\sigma)} f\right)_{\mathbb{R}^N} d\sigma.$$

[3] This is, by no means, the most efficient way to arrive at Bochner's identity. On the other hand, it is amusing and demonstrates the connection with what we have been doing.

Hence

$$\Delta_M f(q(s)) - \Delta_M f(x) = 2\lim_{t\searrow 0}\int_0^s \frac{U(t,q(\sigma)) - U(0,q(\sigma))}{t}\,d\sigma$$
$$- \int_0^s \left(\mathrm{Ric}_{q(\sigma)}\Pi^M_{q(\sigma)}\boldsymbol{\xi}, \mathrm{grad}^M_{q(\sigma)} f\right)_{\mathbb{R}^N} d\sigma,$$

and therefore

$$D^M_{\boldsymbol{\xi}} \circ \Delta_M f(x) = 2\dot{U}(0,x) - \left(\mathrm{Ric}_x\Pi^M_x\boldsymbol{\xi}, \mathrm{grad}^M_x f\right)_{\mathbb{R}^N}.$$

To complete the proof from here, all that we have to show is that

$$2\dot{U}(0,x) = \sum_{i=1}^N \left(\Pi^M_x\boldsymbol{\xi}, \left((\nabla^M_{D^{\mathfrak{e},M}_i})^2 \mathrm{grad}^M f\right)_x\right)_{\mathbb{R}^N}.$$

But, by Theorem 5.8,

$$2\dot{U}(0,x) = \sum_{i=1}^N (\mathfrak{D}^{\mathfrak{e},M}_i)_{(x,I)} \circ \mathfrak{D}^{\mathfrak{e},M}_i \psi_x$$

where

$$\psi_x(y,O) = \left(O\Pi^M_x\boldsymbol{\xi}, \mathrm{grad}^M_y f\right)_{\mathbb{R}^N};$$

and, by two applications of (6.10),

$$(\mathfrak{D}^{\mathfrak{e},M}_i)_{(x,I)} \circ \mathfrak{D}^{\mathfrak{e},M}_i \psi_x = \left(\Pi^M_x\boldsymbol{\xi}, \left((\nabla^M_{D^{\mathfrak{e},M}_i})^2 \mathrm{grad}^M f\right)_x\right)_{\mathbb{R}^N}. \quad \square$$

6.1.4. The Case of Positive Ricci Curvature. Bochner's identity is the starting point for a remarkably large fraction of the known results in which analytic quantities are estimated in terms of geometric ones.[4] The example given here provides a particularly elementary demonstration of this point.

Suppose that (5.46) holds for some $\gamma \in \mathbb{R}$. Then

$$\text{(6.20)} \qquad \|\hat{J}_{\mathfrak{p}}(t)\|_{\mathrm{op}} \le e^{-\gamma t} \quad \text{for any } (t,\mathfrak{p}) \in [0,\infty) \times \mathcal{P}(M \times O(\mathbb{R}^N)).$$

To check this, let $\boldsymbol{\xi} \in \mathbb{R}^N$ be given, and set $\boldsymbol{\xi}(t) = \hat{J}_{\mathfrak{p}}(t)\boldsymbol{\xi}$. Then, because $\mathrm{Ric}_x : T_xM \longrightarrow T_xM$,

$$\frac{d}{dt}|\boldsymbol{\xi}(t)|^2_{\mathbb{R}^N} = 2(\dot{\boldsymbol{\xi}}(t), \boldsymbol{\xi}(t))_{\mathbb{R}^N} = -\left(\mathrm{Ric}_{p(t)}\Pi^M_{p(t)}O(t)\boldsymbol{\xi}(t), O(t)\boldsymbol{\xi}(t)\right)_{\mathbb{R}^N}$$
$$\le -2\gamma|\boldsymbol{\xi}(t)|^2_{\mathbb{R}^N},$$

[4] The grand master of such applications is S.-T. Yau. In particular, Yau and his school (cf. [27]) are responsible for establishing Harnack inequalities on the basis of lower bounds on Ricci curvature, all of which corroborate, in one way or another, the principle, enunciated earlier, that lower bounds on the Ricci curvature control growth rates of Brownian motion.

from which the desired estimate is obvious.

Our goal in this subsection is to use (6.20) to show how $\gamma > 0$ leads to interesting information about the longtime behavior of Brownian paths. In order to put these considerations in context, recall that in § 5.3.2, especially in (5.47) and (5.48), we saw that a uniform lower bound on the Ricci curvature gives strong control on the rate at which Brownian motion can spread out. However, when the γ in (5.46) is strictly positive, the conclusions drawn there are interesting only for short time. In fact, by a famous theorem of Myers (cf. Theorem 3.1 in Chapter 10 of [8]),

$$\gamma > 0 \implies \operatorname{diam}(M) \leq \pi\sqrt{\frac{\dim(M)-1}{2\gamma}} \tag{6.21}$$

when M is connected. In particular, each connected component of M must be compact, and therefore the Brownian motion has nowhere to escape. Thus, it is reasonable to suppose that it will disperse as best it can under the circumstances: it will attempt to become equi-distributed over M. That this is indeed the case is the content of the following simple application of (6.20) combined with (6.17). (See § 10.3.2 for a more dramatic demonstration.)

6.22 THEOREM. *Assume that M is connected and that (5.46) holds for some $\gamma > 0$. Then M is compact and so $\lambda_M(M) < \infty$. Moreover, if $\bar{\lambda}_M \equiv \lambda_M(M)^{-1}\lambda_M$, then (cf. (6.21))*

$$\sup_{x\in M}\left|[\mathbf{P}_T^M f](x) - \int_M f\,d\bar{\lambda}_M\right| \leq \operatorname{diam}(M)\|\operatorname{grad}^M f\|_{C(M;\mathbb{R}^N)}e^{-\gamma T} \tag{6.23}$$

for all $f \in C^1(M;\mathbb{R})$ and $T \in (0,\infty)$.

PROOF: The first part is a simply a restatement of Myers's Theorem. Knowing that M is compact, we also know that (6.17) holds. Hence, in conjunction with (6.20), we see first that

$$\left|\operatorname{grad}_x^M \mathbf{P}_T^M f\right|_{\mathbb{R}^N} \leq e^{-\gamma T}\|\operatorname{grad}^M f\|_{C(M;\mathbb{R}^N)}$$

and then that

$$\begin{aligned}\left|[\mathbf{P}_T^M f](x) - [\mathbf{P}_T^M f](y)\right| &\leq e^{-\gamma T}\|\operatorname{grad}^M f\|_{C(M;\mathbb{R}^N)}\operatorname{dist}^M(x,y)\\ &\leq e^{-\gamma T}\|\operatorname{grad}^M f\|_{C(M;\mathbb{R}^N)}\operatorname{diam}(M).\end{aligned}$$

After integrating both sides in y with respect to $\bar{\lambda}_M$, we get

$$\left|[\mathbf{P}_T^M f](x) - \int_M \mathbf{P}_T^M f\,d\bar{\lambda}_M\right| \leq e^{-\gamma T}\|\operatorname{grad}^M f\|_{C(M;\mathbb{R}^N)}\operatorname{diam}(M).$$

Finally, by the last part of Theorem 4.37 with $g = \mathbf{1}$,

$$\int_M \mathbf{P}_T^M f\,d\bar{\lambda}_M = \int_M f\,d\bar{\lambda}_M. \quad \square$$

The preceding provides some, albeit crude, information about the spectrum $\mathrm{Spec}(-\Delta_M)$ of $-\Delta_M$ as an operator on $L^2(\lambda_M;\mathbb{R})$. In the first place, it shows that 0 is a simple eigenvalue whose eigenfunctions are constant. Indeed, $\mathbf{1}$ is certainly a non-zero eigenfunction with eigenvalue 0. Moreover, if f is any eigenfunction with this eigenvalue, then[5] $f = \mathbf{P}_T^M f$ for all $T \in (0,\infty)$, and therefore, by (6.23), $f \perp \mathbf{1}$ in $L^2(\lambda_M,\mathbb{R})$ implies that $f = \mathbf{0}$. Next, let $\alpha \in \mathrm{Spec}(-\Delta_M)\setminus\{0\}$. If f is an associated, non-zero eigenfunction, then $f \perp \mathbf{1}$ and $\mathbf{P}_T^M f = e^{-\frac{\alpha}{2}T} f$ for all $T \in (0,\infty)$. Hence, (6.23) now leads to the conclusion that $\mathrm{Spec}(-\Delta_M) \setminus \{0\} \subseteq [2\gamma,\infty)$.

Actually, one can improve this estimate by starting from (6.19) instead of (6.17). Namely, again take f to be an eigenvector associated with a non-zero eigenvalue α, only this time assume that $\|f\|_{L^2(\lambda_M;\mathbb{R})} = 1$. We already know that $\alpha > 0$. To do better, we use (6.19) to justify:

$$\alpha\,\mathrm{grad}^M f = -\mathrm{grad}^M \Delta_M f = -\sum_{i=1}^N \nabla^2_{D_i^{\mathfrak{e},M}} \mathrm{grad}^M f + \mathrm{Ric}\,\mathrm{grad}^M f,$$

from which we see that $\alpha\left|\mathrm{grad}^M f\right|^2_{\mathbb{R}^N}$ is equal to

$$-\sum_{i=1}^N \Big(\mathrm{grad}^M f, \nabla^2_{D_i^{\mathfrak{e},M}} \mathrm{grad}^M f\Big)_{\mathbb{R}^N} + \Big(\mathrm{grad}^M f, \mathrm{Ric}\,\mathrm{grad}^M f\Big)_{\mathbb{R}^N}.$$

Next, by (5.14),

$$(D_i^{\mathfrak{e},M})^2\left|\mathrm{grad}^M f\right|^2_{\mathbb{R}^N} = 2\Big(\mathrm{grad}^M f, \nabla^2_{D_i^{\mathfrak{e},M}} \mathrm{grad}^M f\Big)_{\mathbb{R}^N} + 2\left|\nabla_{D_i^{\mathfrak{e},M}} \mathrm{grad}^M f\right|^2_{\mathbb{R}^N},$$

and so another expression for $\alpha\left|\mathrm{grad}^M f\right|^2_{\mathbb{R}^N}$ is

$$\sum_{i=1}^N \left|\nabla_{D_i^{\mathfrak{e},M}} \mathrm{grad}^M f\right|^2_{\mathbb{R}^N} - \tfrac{1}{2}\Delta_M\left|\mathrm{grad}^M f\right|^2_{\mathbb{R}^N} + \Big(\mathrm{grad}^M f, \mathrm{Ric}\,\mathrm{grad}^M f\Big)_{\mathbb{R}^N}.$$

Moreover, by (4.30) and the second equality in (5.12), one sees that

$$\left|\nabla_{D_i^{\mathfrak{e},M}} \mathrm{grad}^M f\right|^2_{\mathbb{R}^N} = \sum_{j,j'=1}^N \big(D_i^{\mathfrak{e},M} \circ D_j^{\mathfrak{e},M} f\big)\big(D_i^{\mathfrak{e},M} \circ D_{j'}^{\mathfrak{e},M} f\big)\big(\Pi \mathbf{e}_j, \mathbf{e}_{j'}\big)_{\mathbb{R}^N},$$

and, since all these considerations must be independent of the choice of the orthonormal basis $\mathfrak{e}$, we can, for a given $x \in M$, choose $\mathfrak{e}$ so that $\mathbf{e}_1, \dots, \mathbf{e}_{\dim(M)} \in T_x M$ and thereby achieve

$$\sum_{i=1}^N \left|\nabla_{D_i^{\mathfrak{e},M}} \mathrm{grad}^M f\right|^2_{\mathbb{R}^N} = \sum_{i,j=1}^{\dim(M)} \big(D_i^{\mathfrak{e},M} \circ D_j^{\mathfrak{e},M} f\big)^2$$

$$\geq \sum_{i=1}^{\dim(M)} \big((D_i^{\mathfrak{e},M})^2 f\big)^2 \geq \frac{1}{\dim(M)}\big(\Delta_M f\big)^2 = \frac{\alpha^2}{\dim(M)} f^2$$

[5] We are tacitly using basic elliptic regularity results which guarantee that the spectrum of $-\Delta_M$ consists entirely of eigenvalues and that the eigenfunctions are smooth.

at that x. Hence, after combining the above, we obtain

$$\alpha\big|\mathrm{grad}^M f\big|^2_{\mathbb{R}^N} \geq \frac{\alpha^2}{\dim(M)} f^2 + 2\gamma\big|\mathrm{grad}^M f\big|^2_{\mathbb{R}^N} - \tfrac{1}{2}\Delta_M\big|\mathrm{grad}^M f\big|^2_{\mathbb{R}^N}.$$

Finally, because

$$\int_M \Delta_M\big|\mathrm{grad}^M f\big|^2_{\mathbb{R}^N}\, d\lambda_M = 0$$

and

$$\int_M \big|\mathrm{grad}^M f\big|^2_{\mathbb{R}^N}\, d\lambda_M = -\int_M f\Delta_M f\, d\lambda_M = \alpha\int_M f^2\, d\lambda_M = \alpha,$$

integration of the preceding inequality over M brings us to

$$\alpha^2 \geq \frac{\alpha^2}{\dim(M)} + 2\gamma\alpha.$$

In other words,

$$\gamma > 0 \implies \mathrm{Spec}(-\Delta_M)\setminus\{0\} \subseteq \left[\frac{2\dim(M)\gamma}{\dim(M)-1}, \infty\right) \tag{6.24}$$

which is an example of a family of estimates due to Lichnerowitz.

6.2 Applications of Bochner's Identity

Throughout this section we will be assuming that M is a closed, connected, embedded submanifold of $\mathbb{R}^N$ on which (5.46) holds for some $\gamma \in (-\infty, 0]$. In § 6.1.2, we proved (6.17) under the assumption that the transformation C_x (cf. (6.12)) is bounded above independent of $x \in M$. On the other hand, (6.20) would incline one to believe that (5.46) alone should be enough to guarantee the validity of (6.17). In the present section we will show how, by combining (6.19) with a few crucial facts from analysis, one can justify this belief.[6]

6.2.1. A Couple of Important Analytic Facts. There are two critical pieces of analytic input which we require to carry out the program of this section. The first of these is the following standard regularity result about solutions to non-degenerate parabolic equations.

6.25 THEOREM. *There exists a smooth function, known as the heat kernel,* $(t, x, y) \in (0,\infty)\times M\times M \longmapsto g_t(x,y) \in (0,\infty)$ *such that*

$$[\mathbf{P}^M_t f](x) = \int_M f(y) g_t(x,y)\, \lambda_M(dy), \quad f \in C_{\mathrm{b}}(M;\mathbb{R}).$$

[6] Later on, in §10.3, we will an entirely different line of reasoning to show that (6.17) holds as soon as one has (5.46).

In particular, for each $f \in C_c(M;\mathbb{R})$,

$$\partial_t \mathbf{P}_t^M f = \tfrac{1}{2}\Delta_M \mathbf{P}_t^M f \quad \text{on } (0,\infty)\times M,$$

and so $\Delta_M \mathbf{P}_t^M f = \mathbf{P}_t^M \Delta_M f$ *if* $f \in C_c^2(M;\mathbb{R})$*. Moreover, for each* $t > 0$, g_t *is symmetric, in the sense that* $g_t(x,y) = g_t(y,x)$*, and* g *satisfies the Chapman–Kolmogorov equation:*

$$g_{s+t}(x,y)\int_M g_s(x,z)g_t(z,y)\,\lambda_M(dz);$$

and so $\|g_t(x,\,\cdot\,)\|^2_{L^2(\lambda_M;\mathbb{R})} = g_{2t}(x,x) < \infty$ *for all* $(t,x)\in(0,\infty)\times M$.

PROOF: There are many approaches to proving these assertions. However, we will indicate only two of them and will give the details of neither. The first, and the more pedestrian, approach is to use classical analytic techniques to produce a minimal, smooth, positive fundamental solution $(t,x,y)\rightsquigarrow g_t(x,y)$ to the heat equation on $(0,\infty)\times M$, and then use the martingale characterization of $\mathbb{P}_x^M$ to verify that, for each $t>0$, g_t must be the kernel for $\mathbf{P}_t^M$. Although this approach is elementary, it has certain annoying features stemming from the fact that the requisite "classical analytic techniques" (cf. [16] or [25]) are most readily available in the context of uniformly elliptic operators on Euclidean space. Thus, this approach requires a good deal of unpleasant localization via partitions of unity.

A second, and more elegant, approach is to take the maximal advantage of what we already know. That is, start by defining the distribution (in the sense of L. Schwartz) u on $(0,\infty)\times M\times M$ by

$$(u,\psi) = \iiint\limits_{(0,\infty)\times M\times M} \left[\mathbf{P}_t^M\psi(x,\,\cdot\,)\right](x)\,dt\lambda_M(dx)\lambda_M(dy)$$

for all test functions $\psi\in C_c^\infty\big((0,\infty)\times M\times M\big)$, and check that u is a "weak solution" to the heat equation on $(0,\infty)\times M\times M$ in the sense that, for any test function ψ, $(u,\psi')=0$ when

$$\psi'(t,x,y)\equiv\partial_t\psi(t,x,y)+\frac{1}{4}\big[\Delta_M\psi(t,\,\cdot\,,y)\big](x)+\frac{1}{4}\big[\Delta_M\psi(t,x,\,\cdot\,)\big](y).$$

To verify this fact, one should first check that it suffices to handle ψ's of the form $\psi_0(t)\psi_1(x)\psi_2(y)$. But

$$\begin{aligned}
&\int_{(0,\infty)}\partial_t\psi_0(t)\left(\int_M\psi_1(x)\big[\mathbf{P}_t^M\psi_2\big](x)\,\lambda_M(dx)\right)dt\\
&\qquad=-\int_{(0,\infty)}\psi_0(t)\left(\int_M\psi_1(x)\big[\mathbf{P}_t^M\tfrac{1}{2}\Delta_M\psi_2\big](x)\,\lambda_M(dx)\right)dt\\
&\qquad=-\int_{(0,\infty)}\psi_0(t)\left(\int_M\psi_2(y)\big[\mathbf{P}_t^M\tfrac{1}{2}\Delta_M\psi_1\big](y)\,\lambda_M(dx)\right)dt,
\end{aligned}$$

where we have used integration by parts and $\partial_t \mathbf{P}_t^M \psi_1 = \mathbf{P}_t^M \frac{1}{2}\Delta_M \psi_1$ to get the first equality and symmetry (cf. the last part of Theorem 4.37) of $\mathbf{P}_t^M$ in $L^2(\lambda_M;\mathbb{R})$ to get the second. Hence, $(u,\psi')=0$ is easy for such ψ's, and is therefore true in general. But knowing that u is a weak solution to the heat equation on $(0,\infty)\times M\times M$ allows us to apply well established *hypoelliptic* regularity results for solutions to non-degenerate parabolic equations. Namely, it is known[7] that a weak solution to a non-degenerate parabolic in an open set is necessarily smooth in that open set. Notice that, although this hypoellipticty result is usually proved in the Euclidean context, localization causes no problem here. Indeed, the statement is inherently local!

Whatever method one chooses, once g is at hand, it is clear that g must be a non-zero, non-negative function which satisfies $\partial_t g_t(\,\cdot\,,y) = \frac{1}{2}\Delta_M g_t(\,\cdot\,,y)$ on $(0,\infty)\times M$ for each $y\in M$. Thus, since M is connected, the strong maximum principle for non-degenerate parabolic equations applies and guarantees that g is everywhere strictly positive. Further, the symmetry of g_t follows immediately from the symmetry of $\mathbf{P}_t^M$, and the Chapman–Kolmogorov equation is an equally simple application of $\mathbf{P}_{s+t}^M = \mathbf{P}_s^M \circ \mathbf{P}_t^M$. Hence, after combining these, we have that

$$\|g_t(x,\,\cdot\,)\|_{L^2(\lambda_M;\mathbb{R})}^2 = \int_M g_t(x,z)g_t(z,x)\,\lambda_M(dz) = g_{2t}(x,x).$$

To complete the proof from here, all that we have to do is check that for any $f\in C_c(M;\mathbb{R})$, $t\in(0,\infty)\longmapsto u(t)\equiv \mathbf{P}_t^M f\in C^\infty(M;\mathbb{R})$ is a smooth map which satisfies $\partial_t u = \frac{1}{2}\Delta_M u$. To this end, first note that there is no question about the smoothness. Thus, what remains is to verify $\partial_t u(t) = \frac{1}{2}\Delta_M u(t)$. For this purpose, let $x\in M$ be given, choose $\eta\in C_c^\infty\big(M;[0,1]\big)$ so that $\eta=1$ in an open neighborhood U of x, and set $v(t)=\eta u(t)$. Then

$$\begin{aligned} u(t+h,x)-u(t,x) &= \big[\mathbf{P}_h^M u(t)\big](x) - u(t,x) \\ &= \big[\mathbf{P}_h^M v(t)\big](x) - v(t,x) + \big[\mathbf{P}_h^M (1-\eta)u(t)\big](x). \end{aligned}$$

Since $v(t)\in C_c^\infty(M;\mathbb{R})$,

$$\begin{aligned} \frac{\big[\mathbf{P}_h^M v(t)\big](x) - v(t,x)}{h} &= \frac{1}{h}\int_0^h \big[\mathbf{P}_\tau^M \tfrac{1}{2}\Delta_M v(t,\,\cdot\,)\big](x)\,d\tau \\ &\longrightarrow \big[\tfrac{1}{2}\Delta_M v(t,\,\cdot\,)\big](x) = \big[\tfrac{1}{2}\Delta_M u(t,\,\cdot\,)\big](x). \end{aligned}$$

On the other hand,

$$\big|\big[\mathbf{P}_h^M (1-\eta)u(t)\big](x)\big| \le \|f\|_{C_b(M;\mathbb{R})}\mathbb{P}_x^M\big(p(h)\notin U\big),$$

and so (5.47) is much more than enough to guarantee that, as $h\searrow 0$, this term goes to 0 at least as fast as $e^{-\frac{\epsilon}{h}}$ for some $\epsilon>0$. □

The second fact which we will use is less standard and includes a beautiful inequality proved by Li and Yau in Theorem 1.3 of [27].

[7] This is a special case of Hörmander's famous hypoellipticity theorem in the paper [21] alluded to in connection with (2.3). Alternatively, one can look at Chapter XV of [43], which is devoted specifically to the parabolic case.

6.26 LEMMA. *Assume that* $f \in C_c^\infty(M;[0,\infty))$. *Then there exists a* $C \in (0,\infty)$, *depending only on* d, *such that, for each* $t \in (0,\infty)$,

$$
\begin{aligned}
&\left|\operatorname{grad}_x^M \mathbf{P}_t^M f\right|_{\mathbb{R}^N}^2 \\
&\quad \le 3\left[\mathbf{P}_t^M f\right](x)\left[\mathbf{P}_t^M \tfrac{1}{2}\Delta_M f\right](x) + \left(\frac{9\dim(M)}{4t} + C\right)\left[\mathbf{P}_t^M f\right](x)^2.
\end{aligned}
\tag{6.27}
$$

In addition,

$$
\left\|\operatorname{grad}^M \mathbf{P}_t^M f\right\|_{L^2(\lambda_M;\mathbb{R}^N)}^2 \le \|f\|_{L^2(\lambda_M;\mathbb{R})}\|\Delta_M f\|_{L^2(\lambda_M;\mathbb{R})},
$$

and, as $t \searrow 0$, $\operatorname{grad}^M \mathbf{P}_t^M f \longrightarrow \operatorname{grad}^M f$ *in* $L^2(\lambda_M;\mathbb{R}^N)$.

PROOF: Set $u(t) = \mathbf{P}_t^M f$ and $\mathbf{U}(t) = \operatorname{grad}^M u(t)$. Then the inequality of Li and Yau states (cf. Theorem 1.2 in [27], and take $\alpha = \frac{3}{2}$ there) that

$$
\frac{|\mathbf{U}(2t)|_{\mathbb{R}^N}^2}{u(2t)^2} - \frac{3}{2}\frac{\Delta_M u(2t)}{u(2t)} \le \frac{9\dim(M)}{8t} + C
$$

for a C with the stated dependence. Hence (6.27) follows from the fact that $\frac{1}{2}\Delta_M u(t) = \mathbf{P}_t^M \frac{1}{2}\Delta_M f$.

To prove the rest of the lemma, we begin by recalling (cf. (3.55)) that $\mathbf{P}_t^M$ is a contraction on $L^2(\lambda_M;\mathbb{R})$ for each $t > 0$. Next, choose $\eta \in C_c^\infty(\mathbb{R}^N;[0,1])$ so that $\eta \equiv 1$ in the ball of radius 1 around the origin and the Euclidean gradient of η has length dominated by 1 everywhere, and set $\eta_R(x) = \eta(R^{-1}x)$ for $R \ge 1$. It is then clear that $|\operatorname{grad}^M \eta_R|_{\mathbb{R}^N} \le R^{-1}$ everywhere on M. Moreover, elementary integration by parts followed by Schwarz's inequality leads to:

$$
\begin{aligned}
&\int_M \eta_R^2 |\mathbf{U}(t)|_{\mathbb{R}^N}^2 \, d\lambda_M \\
&\quad = -\int_M \Big(2\eta_R u(t)\big(\operatorname{grad}^M \eta_R, \mathbf{U}(t)\big)_{\mathbb{R}^N} + \eta_R^2 u(t)\Delta_M u(t)\Big) d\lambda_M \\
&\quad \le \frac{2}{R}\|u(t)\|_{L^2(\lambda_M;\mathbb{R})}\big\|\eta_R \mathbf{U}(t)\big\|_{L^2(\lambda_M;\mathbb{R}^N)} + \|f\|_{L^2(\lambda_M;\mathbb{R})}\|\Delta_M f\|_{L^2(\lambda_M;\mathbb{R})},
\end{aligned}
$$

where, in the last line, we have used $\Delta_M u(t) = \mathbf{P}_t^M \Delta_M f$ and the L^2-contraction property of $\mathbf{P}_t^M$. Starting from the preceding, it is an elementary exercise to show first that the left hand side in the first line is bounded independent of $R \ge 1$ and then that

$$
\begin{aligned}
\int_M |\mathbf{U}(t)|_{\mathbb{R}^N}^2 \, d\lambda_M &= -\int_M u(t)\mathbf{P}_t^M \Delta_M f \, d\lambda_M = -\int_M u(2t)\Delta_M f \, d\lambda_M \\
&\longrightarrow -\int_M f\Delta_M f \, d\lambda_M \quad \text{as } t \searrow 0.
\end{aligned}
$$

In particular, by Schwarz's inequality, the second equality proves the asserted estimate for $\|\mathbf{U}(t)\|_{L^2(\lambda_M;\mathbb{R}^N)}$. In addition, we also see that if $\mathbf{F} = \mathrm{grad}^M f$, then $\|\mathbf{U}(t) - \mathbf{F}\|^2_{L^2(\lambda_M;\mathbb{R}^N)}$ equals

$$\begin{aligned}&\|\mathbf{U}(t)\|^2_{L^2(\lambda_M;\mathbb{R}^N)} - 2\big(\mathbf{U}(t),\mathbf{F}\big)_{L^2(\lambda_M;\mathbb{R}^N)} + \|\mathbf{F}\|^2_{L^2(\lambda_M;\mathbb{R}^N)}\\ &\quad = \|\mathbf{U}(t)\|^2_{L^2(\lambda_M;\mathbb{R}^N)} + 2\big(u(t),\Delta_M f\big)_{L^2(\lambda_M;\mathbb{R})} - \big(f,\Delta_M f\big)_{L^2(\lambda_M;\mathbb{R})} \longrightarrow 0. \quad \square\end{aligned}$$

6.2.2. Integrating Bochner's Identity. We can now show how to reverse the differentiation process by which we passed from (6.17) to (6.19). That is, we are here going to integrate (6.19) and thereby obtain (6.17) under the condition (5.46).

6.28 THEOREM. *Suppose that $f \in C^1(M;\mathbb{R})$ with $|\mathrm{grad}^M f|_{\mathbb{R}^N}$ bounded. Then $\mathbf{P}^M_t f$ exists (i.e., $f(p(t))$ is $\mathbb{P}^M_x$-integrable) and is an element of $C^1(M;\mathbb{R})$ for each $t > 0$. In fact, (cf. Theorem 5.8 and (6.14)) (6.17) holds. Equivalently, for each $(T,x) \in (0,\infty) \times M$,*

$$\mathrm{grad}^M_x \mathbf{P}^M_T f = \mathbb{E}^{\widetilde{\mathbb{P}^M_x}}\Big[\big(O(T)\hat{J}_{\mathrm{p}}(T)\big)^\top \mathrm{grad}^M_{p(t)} f\Big]. \tag{6.29}$$

In particular, (cf. (6.20))

$$\big|\mathrm{grad}^M \mathbf{P}^M_T f\big|_{\mathbb{R}^N} \le e^{-\gamma T}\mathbf{P}^M_T \big|\mathrm{grad}^M f\big|_{\mathbb{R}^N}. \tag{6.30}$$

PROOF: We begin with the case when $f \in C^\infty_c(M;[0,\infty))$, in which case Theorem 6.25 guarantees that u is smooth and Lemma 6.26 applies. Set $\mathbf{F} = \mathrm{grad}^M f$, $u(t) = \mathbf{P}^M_t f$, and $\mathbf{U}(t) = \mathrm{grad}^M u(t)$. Given $1 \le k \le N$ and $\epsilon > 0$, define $\Phi_{k,\epsilon} : [0,T] \times M \times O(\mathbb{R}^N) \longrightarrow \mathbb{R}$ by

$$\Phi_{k,\epsilon}(t,x,O) = \big(\mathbf{e}_k, O^\top U(T+\epsilon - t\wedge T, x)\big)_{\mathbb{R}^N}.$$

Then (cf. (5.7)), by (6.10), for $t \in (0,T)$,

$$\begin{aligned}&\frac{\partial \Phi_{k,\epsilon}}{\partial t}(t,x,O) + \frac{1}{2}\sum_{i=1}^N (\mathfrak{D}^{\epsilon,M}_i)_{(x,O)} \circ \mathfrak{D}^{\epsilon,M}_i \Phi_{k,\epsilon}(t)\\ &\qquad = -\Big(\mathbf{e}_k, O^\top \mathrm{grad}^M_x \tfrac{1}{2}\Delta_M u(t)\Big)_{\mathbb{R}^N}\\ &\qquad\qquad + \left(\mathbf{e}_k, O^\top \frac{1}{2}\sum_{i=1}^N (\nabla_{D^{\epsilon,M}_i})_x \circ \nabla_{D^{\epsilon,M}_i}\mathbf{U}(T+\epsilon - t)\right)_{\mathbb{R}^N}\\ &\qquad = \frac{1}{2}\Big(\mathbf{e}_k, O^\top \mathrm{Ric}_x \mathbf{U}(T+\epsilon - t, x)\Big)_{\mathbb{R}^N},\end{aligned}$$

where we have used (6.19) in the passage to the last line. Hence, by our standard cut-off procedure plus Doob's Stopping Time Theorem, we know that, for each $R > 0$,

$$\begin{aligned}X_{k,\epsilon,R}(t,\mathrm{p}) &\equiv \Phi_{k,\epsilon}\big(t\wedge\zeta_R, \mathrm{p}(t\wedge T\wedge\zeta_R)\big)\\ &\qquad - \int_0^{t\wedge T\wedge\zeta_R(\mathrm{p})} \tfrac{1}{2}\Big(\mathbf{e}_k, O(\tau)^\top \mathrm{Ric}_{p(\tau)}\mathbf{U}\big(T+\epsilon-\tau, p(\tau)\big)\Big)_{\mathbb{R}^N}\, d\tau\end{aligned}$$

is a $\widetilde{\mathbb{P}^M_x}$-martingale when $\zeta_R(p)$ is the first time that $p \in \mathbb{P}(M)$ goes a distance R from its starting point. Thus, if

$$\mathbf{Y}_k(t,\mathrm{p}) \equiv \hat{J}_{\mathrm{p}}(t \wedge T \wedge \zeta_R)^\top \mathbf{e}_k,$$

then a trivial vector valued extention of Lemma 2.41 shows that

$$\begin{aligned}\big(O(t \wedge T \wedge \zeta_R)\hat{J}_{\mathrm{p}}(t \wedge T \wedge \zeta_R)\big)^\top \mathbf{U}\big(T + \epsilon - t \wedge T \wedge \zeta_R, \mathrm{p}(t \wedge T \wedge \zeta_R)\big) \\ - \int_0^{t \wedge T \wedge \zeta_R} \big(O(\tau)\hat{J}_{\mathrm{p}}(\tau)\big)^\top \mathbf{U}\big(T + \epsilon - \tau, p(\tau)\big)\, d\tau \\ = \sum_{k=1}^N \left(X_{k,\epsilon}(t,\mathrm{p})\mathbf{Y}_k(t,\mathrm{p}) - \int_0^t X_{k,\epsilon}(\tau,\mathrm{p})\dot{\mathbf{Y}}_k(\tau,\mathrm{p})\, d\tau \right)\end{aligned}$$

is a $\widetilde{\mathbb{P}^M_x}$-martingale. But, by (6.27), $|\mathbf{U}(T+\epsilon-t,y)|_{\mathbb{R}^N}$ is uniformly bounded[8] as (t,y) runs over $[0,T] \times M$, and so we can now let $R \to \infty$ and say that

$$\big(O(t \wedge T)\hat{J}_{\mathrm{p}}(t \wedge T)\big)^\top \mathbf{U}\big(T + \epsilon - t \wedge T, \mathrm{p}(t \wedge T)\big)$$

is a $\widetilde{\mathbb{P}_x}$-martingale. In particular, this leads to

$$\mathbf{U}(T + \epsilon, x) = \mathbb{E}^{\widetilde{\mathbb{P}_x}}\Big[\big(O(T)\hat{J}_{\mathrm{p}}(T)\big)^\top \mathbf{U}\big(\epsilon, \mathrm{p}(T)\big)\Big].$$

But the difference between the right hand side of the preceding and the right hand side of (6.29) is dominated by (cf. Theorem 6.25)

$$\begin{aligned}\mathbb{E}^{\mathbb{P}^M_x}\Big[\big|\mathbf{U}(\epsilon, p(T)) - \mathbf{F}\big(p(T)\big)\big|_{\mathbb{R}^N}\Big] &= \int_M g_T(x,y)\big|\mathbf{U}(\epsilon,y) - \mathbf{F}(y)\big|_{\mathbb{R}^N} \\ &\le g_{2T}(x,x)^{\frac{1}{2}} \|\mathbf{U}(\epsilon,\,\cdot\,) - \mathbf{F}\|_{L^2(\lambda_M;\mathbb{R}^N)},\end{aligned}$$

which, by the last part of Lemma 6.26, tends to 0 as $\epsilon \searrow 0$. In other words, we have now proved (6.29) for non-negative $f \in C^\infty_{\mathrm{c}}(M;\mathbb{R})$.

Dropping the non-negativity assumption is trivial. Namely, if $f \in C^\infty_{\mathrm{c}}(M;\mathbb{R})$, then we can certainly write f as $f_1 - f_2$, where both f_1 and f_2 are non-negative elements of $C^\infty_{\mathrm{c}}(M;\mathbb{R})$. Hence, by linearity, (6.29) holds for such f's. To handle $f \in C^1_{\mathrm{c}}(M;\mathbb{R})$, recall that we can extend f as a compactly supported element $\bar{f} \in C^1(\mathbb{R}^N;\mathbb{R})$, and so standard mollification procedures applied to $\bar{f}$ lead to our having a sequence $\{f_n\}_1^\infty \subseteq C^\infty_{\mathrm{c}}(M;\mathbb{R})$ with the properties that they all vanish off of a fixed compact and $f_n \to f$ in $C^1_{\mathrm{b}}(M;\mathbb{R})$. Hence, (6.29) extends to f by passing to the limit in (6.29) as $n \to \infty$ after replacing f by f_n

[8] Because this qualitative conclusion is our only application of (6.27), it seems likely that one can get away without using the Li–Yau sledge hammer to kill this fly. As we already mentioned, Chapter 10 provides such an alternative, but the contents of Chapter 10 are *not* exactly elementary.

on both sides. Indeed, it is obvious that the right hand side converges to the desired quantity uniformly in $x \in M$. At the same time, since $\mathbf{P}_T^M f_n$ tends to $\mathbf{P}_T^M f$ uniformly on M, it is easy to conclude that $\mathbf{P}_t^M f \in C_{\mathrm{b}}^1(M;\mathbb{R})$ and that (6.29) holds.

To complete the proof, let $f \in C^1(M;\mathbb{R})$ with bounded gradient be given, take the same cut-off functions $\{\eta_R : R \geq 1\}$ as in the proof of Lemma 6.26. and set $f_R = \eta_R f$. First note that because $f(y)$ can tend to infinity no faster than $\mathrm{dist}^M(x,y)$, (5.48) assures us that $f\big(p(t)\big)$ is $\mathbb{P}_x^M$-integrable and that

$$\big[\mathbf{P}_t^M f_R\big](x) \longrightarrow \big[\mathbf{P}_t^M f\big](x) \equiv \mathbb{E}^{\mathbb{P}_x^M}\left[f\big(p(t)\big)\right]$$

uniformly on compact subsets of $[0,\infty) \times M$. Hence all that remains is to observe that $\mathrm{grad}^M f_R \longrightarrow \mathrm{grad}^M f$ boundedly on M. $\square$

6.2.3. A Logarithmic Sobolev Inequality. Our goal of this subsection is to show how to go from (6.30) to a very mild form of the Sobolev inequality, known as a *logarithmic Sobolev inequality*, and our derivation will follow a line of reasoning developed by D. Bakry [1].

6.31 THEOREM. *For any* $f \in C^1(M;\mathbb{R}) \setminus \{\mathbf{0}\}$,[9]

$$\left[\mathbf{P}_T^M\left(f^2 \log \frac{f^2}{[\mathbf{P}_T^M f^2](x)}\right)\right](x) \leq \bar{e}_\gamma(2T)\left[\mathbf{P}_T^M\left(\left|\mathrm{grad}^M f\right|_{\mathbb{R}^N}^2\right)\right](x) \tag{6.32}$$

where

$$\bar{e}_\gamma(T) \equiv \int_0^T e^{-\gamma t}\,dt = \frac{1-e^{-\gamma T}}{\gamma} \;(\equiv T \text{ when } \gamma = 0).$$

PROOF: We begin by proving (6.32) for $f \in C^\infty\big(M;(0,\infty)\big)$ which are constant off of some compact. Given such an f, set $\mathbf{F} = \mathrm{grad}^M f^2$, $u(t) = \mathbf{P}_t^M f^2$, $\mathbf{U}(t) = \mathrm{grad}^M u(t)$, and $\ell(t) = u(t)\log u(t)$. By Theorem 6.25, we know that u is smooth. Hence, ℓ is also smooth. In fact, $\Delta_M u(t) = \mathbf{P}_t^M \Delta_M f^2$,

$$\tfrac{1}{2}\Delta_M \ell_t = \frac{|\mathbf{U}(t)|_{\mathbb{R}^N}^2}{2u(t)} + \tfrac{1}{2}\Delta_M u(t)\big(1+\log u(t)\big),$$

and

$$\dot{\ell}_t \equiv \frac{d}{dt}\ell_t = \tfrac{1}{2}\Delta_M u(t)\big(1+\log u(t)\big).$$

In particular, both $t \rightsquigarrow \frac{1}{2}\Delta_M \ell_t$ and $t \rightsquigarrow \dot{\ell}_t$ are continuous maps into $C_{\mathrm{b}}(M;\mathbb{R})$. We now argue that

$$2\frac{d}{dt}\mathbf{P}_t^M \ell_{T-t} = \mathbf{P}_t^M\left(\frac{|\mathbf{U}(T-t)|_{\mathbb{R}^N}^2}{u(T-t)}\right), \quad t \in (0,T), \tag{*}$$

[9] We take $a \log a = 0$ when $a = 0$ and note that then $a \log a \geq -e^{-1}$ for all $a \geq 0$.

which, in view of the preceding calculations, comes down to checking that $\frac{d}{dt}\mathbf{P}_t^M \ell(\tau) = \mathbf{P}_t^M \frac{1}{2}\Delta_M \ell(\tau)$ when $\tau = T - t$. But we have already noted that $\ell(\tau) \in C_b^2(M;\mathbb{R})$, and so this equality is an easy consequence of the martingale characterization of $\mathbb{P}_x^M$.

After combining (*) with (6.30), we see that

$$2\frac{d}{dt}\mathbf{P}_t^M \ell_{T-t} \le e^{-2\gamma(T-t)}\mathbf{P}_t^M\left(\frac{\left(\mathbf{P}_{T-t}^M|\mathbf{F}|_{\mathbb{R}^N}^2\right)^2}{u(T-t)}\right).$$

But, by Schwarz's inequality,

$$\left(\mathbf{P}_{T-t}^M|\mathbf{F}|_{\mathbb{R}^N}^2\right)^2 = 4\left(\mathbf{P}_{T-t}^M f|\mathrm{grad}^M f|_{\mathbb{R}^N}\right)^2 \le 4u(T-t)\mathbf{P}_{T-t}^M|\mathrm{grad}^M f|_{\mathbb{R}^N}^2.$$

Hence, (6.32) results after one plugs this into the preceding and integrates over $[0,T]$.

Having proved it for positive $f \in C^\infty(M;\mathbb{R})$ which are constant off of a compact, one can prove it for any $f \in C^\infty(M;\mathbb{R})$ which is constant off a compact by considering $f_\epsilon = \sqrt{\epsilon^2 + f^2}$ and letting $\epsilon \searrow 0$; and once one has it for these, the result for general f's follows by the same sort of reasoning as we used at the end of the proof of Theorem 6.28. □

Logarithmic Sobolev inequalities are of greatest interest only in extreme situations. For example, as is apparent in (6.32), they, in contrast to bona-fide Sobolev inequalities, are reasonably insensitive to dimension. To give a hint how this fact might be useful, we will need the following observation due to I. Herbst.

6.33 LEMMA. *Suppose that ν is a probability measure on M with the property that, for some $C \in (0,\infty)$,*

$$\text{(6.34)}\qquad \mathbb{E}^\nu\left[f^2\log\frac{f^2}{\|f\|_{L^2(\nu;\mathbb{R})}^2}\right] \le C\mathbb{E}^\nu\left[|\mathrm{grad}^M f|_{\mathbb{R}^N}^2\right]$$

for all $f \in C^\infty(M;(0,\infty))$. If $f \in C^1(M;\mathbb{R})$ and $|\mathrm{grad}^M f|_{\mathbb{R}^N} \le 1$ everywhere, then, for each $0 < \epsilon < C^{-1}$,

$$\text{(6.35)}\qquad \mathbb{E}^\nu\left[\exp(\epsilon f^2)\right] \le \exp\left(\frac{\epsilon\|f\|_{L^2(\nu;\mathbb{R})}^2}{(1-C\epsilon)}\right).$$

PROOF: First observe that, again by the reasoning given at the end of the proof of Theorem 6.28, it suffices for us to handle $f \in C_c^\infty(M;\mathbb{R})$ with $|\mathrm{grad}^M f|_{\mathbb{R}^N} \le 1$. Hence, let such an f be given, and define

$$u(\xi) = \mathbb{E}^\nu\left[e^{2\xi f^2}\right]^{\frac{1}{\xi}} \quad \text{for } \xi \in (0,\infty).$$

Then, by elementary calculus and (6.34) applied to $e^{\xi f^2}$,

$$\begin{aligned} u'(\xi) &= -\xi^{-2}u(\xi)\log \mathbb{E}^{\nu}\left[e^{2\xi f^2}\right] + 2\xi^{-1}u(\xi)^{1-\xi}\mathbb{E}^{\nu}\left[f^2 e^{2\xi f^2}\right] \\ &= \xi^{-2}u(\xi)^{1-\xi}\mathbb{E}^{\nu}\left[e^{2\xi f^2}\log\frac{e^{2\xi f^2}}{\|e^{\xi f^2}\|^2_{L^2(\nu;\mathbb{R})}}\right] \le 4Cu(\xi)^{1-\xi}\mathbb{E}^{\nu}\left[f^2 e^{2\xi f^2}\right] \\ &= 2C\xi u'(\xi) + 2Cu(\xi)\log u(\xi), \end{aligned}$$

and so

$$\frac{d}{d\xi}\log u(\xi) \le \frac{2C\log u(\xi)}{1-2C\xi} \quad \text{for } 0<\xi<(2C)^{-1}$$

Thus, since $\log u(\xi) \longrightarrow 2\|f\|^2_{L^2(\nu;\mathbb{R})}$ as $\xi \searrow 0$, we have proved that

$$\log u(\xi) \le 2\|f\|^2_{L^2(\nu;\mathbb{R})}(1-2C\xi)^{-1}$$

and therefore

$$u(\xi) \le \exp\left(\frac{2\|f\|^2_{L^2(\nu;\mathbb{R})}}{1-2C\xi}\right),$$

which is clearly equivalent to (6.35). □

By combining (6.32) with (6.35), we see that (5.46) implies that

$$\text{(6.36)} \qquad \mathbb{E}^{\mathbb{P}^M_x}\left[\exp\left(\frac{\epsilon f^2(p(T))}{G^2\bar{e}_\gamma(2T)}\right)\right] \le \exp\left(\frac{[\mathbb{P}^M_T f^2](x)}{G^2\bar{e}_\gamma(2T)(1-\epsilon)}\right)$$

for any $f \in C^1(M;\mathbb{R})$ satisfying $\left|\operatorname{grad}^M f\right|_{\mathbb{R}^N} \le G$. Obviously, an estimate on the left hand side of (6.36) can be obtained as an application of (5.48). On the other hand, such an estimate would depend on $\dim(M)$, whereas (6.36) does not.

6.3 Bismut's Formula

The purpose of this section is to show how to pass from (6.17) to an expression in which only f itself, and none of its derivatives, appears on the right. Throughout we will assume that M is a closed, connected, embedded submanifold of $\mathbb{R}^N$ on which (5.46) holds for some $\gamma \in \mathbb{R}$.

6.3.1. Variations on Bochner's Identity. The first step in our program will be to develop a variant of (6.17) via the same technique as we used in the proof of Theorem 6.28.

6.37 LEMMA. *Suppose that $f \in C^1(M;\mathbb{R})$ with $|\operatorname{grad}^M f|_{\mathbb{R}^N}$ bounded. Then (cf. Theorem 5.8 and (6.14)), for each $\boldsymbol{\theta} \in C^1([0,T];\mathbb{R}^N)$ and $T>0$,*

$$\begin{aligned} &\left(D^M_{\boldsymbol{\theta}(T)}\right)_x \mathbf{P}^M_T f - \left(D^M_{\boldsymbol{\theta}(0)}\right)_x \mathbf{P}^M_T f \\ &\qquad = \mathbb{E}^{\widetilde{\mathbb{P}^M_x}}\left[\int_0^T \left(O(t)\hat{J}_{\mathbf{p}}(t)\dot{\boldsymbol{\theta}}(t), \operatorname{grad}^M_{p(t)}\mathbf{P}^M_{T-t}f\right)_{\mathbb{R}^N} dt\right]. \end{aligned}$$

PROOF: Just as in the proof of Theorem 6.28, it suffices to handle $f \in C_c^\infty(M;\mathbb{R})$. Given such an f, set $\mathbf{F} = \mathrm{grad}^M f$, $u(t) = \mathbf{P}_t^M f$, and (cf. Lemma 6.25) $\mathbf{U}(t) = \mathrm{grad}^M u(t)$. Next, let $T \in (0,\infty)$ and $1 \le k \le N$, and define $\Phi_k : [0,\infty) \times M \times O(\mathbb{R}^N) \longrightarrow \mathbb{R}$ given by

$$\Phi_k(t,x,O) = \big(O\mathbf{e}_k, \mathbf{U}(T - t \wedge T, x)\big)_{\mathbb{R}^N}.$$

Then, just as in the proof of Theorem 6.28,

$$\frac{\partial \Phi_k}{\partial t}(t,x,O) + \frac{1}{2}\sum_{i=1}^N (\mathfrak{D}_i^{\mathfrak{t},M})_{(x,O)} \circ \mathfrak{D}_i^{\mathfrak{t},M}\Phi_k(t) = \tfrac{1}{2}\Big(O\mathbf{e}_k, \mathrm{Ric}_x\, \mathbf{U}(T-t,x)\Big)_{\mathbb{R}^N}.$$

Hence (remember that we already know that $\mathbf{U}(t)$ is a bounded continuous function on $[0,T] \times M$)

$$X_k(t,\mathfrak{p}) \equiv \Phi_k\big(t \wedge T, \mathfrak{p}(t \wedge T)\big) - \int_0^{t\wedge T} \tfrac{1}{2}\Big(O(\tau)\mathbf{e}_k, \mathrm{Ric}_{p(\tau)}\big(\mathbf{U}(T-\tau, p(\tau)\big)\Big)_{\mathbb{R}^N} d\tau$$

is a $\widetilde{\mathbb{P}_x^M}$-martingale for any $x \in M$. Finally, take

$$Y_k(t,\mathfrak{p}) = \Big(\hat{J}_{\mathfrak{p}}(t)\boldsymbol{\theta}(t), \mathbf{e}_k\Big)_{\mathbb{R}^N}.$$

Then, by Lemma 2.41,

$$\begin{aligned}
&\Big(O(t \wedge T)\hat{J}_{\mathfrak{p}}(t\wedge T)\boldsymbol{\theta}(t\wedge T), \mathbf{U}\big(T - t\wedge T, p(t\wedge T)\big)\Big)_{\mathbb{R}^N} \\
&\qquad - \int_0^{t\wedge T}\Big(O(\tau)\hat{J}_{\mathfrak{p}}(\tau)\dot{\boldsymbol{\theta}}(\tau), \mathbf{U}\big(T-\tau, p(\tau)\big)\Big)_{\mathbb{R}^N} d\tau \\
&= \sum_{k=1}^N \left(X_k(t\wedge T,\mathfrak{p})Y_k(t\wedge T,\mathfrak{p}) - \int_0^{t\wedge T} X_k(\tau,\mathfrak{p})\dot{Y}_k(\tau,\mathfrak{p})\,d\tau\right)
\end{aligned}$$

is a $\widetilde{\mathbb{P}_x^M}$-martingale; and the required identity results from equating the expected values of this martingale at times $t = 0$ and $t = T$ and then applying (6.29). $\square$

6.3.2. The Bismut Factor. In order to carry out the next step, we will need to make a few preparations. In the first place, given a $\boldsymbol{\theta} \in C^2([0,\infty);\mathbb{R}^N)$, set (cf. (3.17))

$$B_{\boldsymbol{\theta}}(t,x,\mathbf{w}_n) = \int_0^t \Big(O_{p(\,\cdot\,,x,\mathbf{w}_n)}(\tau)\hat{J}_{p(\,\cdot\,,x,\mathbf{w}_n)}(\tau)\dot{\boldsymbol{\theta}}(\tau), \dot{\mathbf{w}}_n(\tau)\Big)_{\mathbb{R}^N} d\tau \tag{6.38}$$

for $(t,x,\mathbf{w}) \in [0,\infty) \times M \times \mathfrak{W}(\mathbb{R}^N)$. Then, by the results in § 3.3 and § 3.4.1, we know that there is a $\{\bar{\mathcal{B}}_t : t \ge 0\}$-progressively measurable map

(6.39)

$$(t,\mathbf{w}) \in [0,\infty) \times \mathfrak{W}(\mathbb{R}^d) \longmapsto \begin{bmatrix} p(t,\,*\,,\mathbf{w}) \\ O(t,\,*\,,\mathbf{w}) \\ \hat{J}(t,\,*\,,\mathbf{w}) \\ B_{\boldsymbol{\theta}}(t,\,*\,,\mathbf{w}) \end{bmatrix}$$
$$\in C^\infty\Big(M; M \times O(\mathbb{R}^N) \times \mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N) \times \mathbb{R}\Big)$$

which is continuous in t for each $\mathbf{w}$ and satisfies

$$\begin{bmatrix} p(t,x,\mathbf{w}_n) \\ O_{\mathrm{p}(\,\cdot\,,x,\mathbf{w}_n)}(t) \\ \hat{J}_{\mathrm{p}(\,\cdot\,,x,\mathbf{w}_n)}(t) \\ B_{\boldsymbol{\theta}}(t,x,\mathbf{w}_n) \end{bmatrix} \longrightarrow \begin{bmatrix} p(t,x,\mathbf{w}) \\ O(t,x,\mathbf{w}) \\ \hat{J}(t,x,\mathbf{w}) \\ B_{\boldsymbol{\theta}}(t,x,\mathbf{w}) \end{bmatrix}$$

uniformly on compact subsets of $[0,\infty)\times M$ for $\mu_{\mathbb{R}^N}$-almost every $\mathbf{w}$. In addition, if, for $A\in \mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N)$, $\rho(A)$ denotes the vector field on $\mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N)$ such that

$$\rho(A)_J f = \frac{d}{ds} f\left(e^{sA}J\right)\Big|_{s=0}, \quad J\in \mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N),$$

then, for each $x\in M$, the $\mu_{\mathbb{R}^N}$-distribution of

$$\mathbf{w}\in \mathfrak{W}(\mathbb{R}^N)\longmapsto \begin{bmatrix} p(\,\cdot\,,x,\mathbf{w}) \\ O(\,\cdot\,,x,\mathbf{w}) \\ \hat{J}(\,\cdot\,,x,\mathbf{w}) \\ B_{\boldsymbol{\theta}}(\,\cdot\,,x,\mathbf{w}) \end{bmatrix} \in C\Big([0,\infty);M\times O(\mathbb{R}^N)\times \mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N)\times\mathbb{R}\Big)$$

is uniquely determined by the fact that it solves the martingale problem starting at $(x,I,I,0)\in M\times O(\mathbb{R}^N)\times \mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N)\times\mathbb{R}$ for the time dependent operator

$$\rho\big(O^\top \mathrm{Ric}_y O\big)_J + \frac{1}{2}\sum_{i=1}^N \big(\mathfrak{Y}_i(t)^{\mathfrak{e},M}\big)^2,$$

where (cf. (5.7))

$$\text{(6.40)}\qquad \big(\mathfrak{Y}_i^{\mathfrak{e},M}(t)\big)_{(y,O,J,B)} = \mathfrak{D}_i^{\mathfrak{e},M} + \big(OJ\dot{\boldsymbol{\theta}}(t),\Pi_y^M \mathbf{e}_i\big)_{\mathbb{R}^N}\partial_B \quad \text{for } 1\le i\le N.$$

In particular, of course,

$$\mathbf{w}\in \mathfrak{W}(\mathbb{R}^N)\longmapsto \mathrm{p}(\,\cdot\,,x,\mathbf{w})\equiv \begin{bmatrix} p(\,\cdot\,,x,\mathbf{w}) \\ O(\,\cdot\,,x,\mathbf{w}) \end{bmatrix} \in C\big([0,\infty);M\times O(\mathbb{R}^N)\big)$$

has $\mu_{\mathbb{R}^N}$-distribution (cf. Theorem 5.8) $\widetilde{\mathbb{P}_x^M}$ and $\hat{J}(\,\cdot\,,x,\mathbf{w}) = \hat{J}_{\mathrm{p}(\,\cdot\,,x,\mathbf{w})}$ for $\mu_{\mathbb{R}^N}$-almost every $\mathbf{w}$.

6.41 Theorem. *Let $\boldsymbol{\theta}\in C^2\big([0,\infty);\mathbb{R}^N\big)$ be given, and define the map $(t,x,\mathbf{w})\rightsquigarrow B_{\boldsymbol{\theta}}(t,x,\mathbf{w})$ accordingly, as in (6.39). Then, for any $\alpha\in\mathbb{R}$, (cf. (5.46) and (6.14))*

$$\text{(6.42)}\qquad \mathbb{E}^{\mu_{\mathbb{R}^N}}\Big[e^{\alpha B_{\boldsymbol{\theta}}(T,x,\mathbf{w})}\Big] \le \exp\left(\frac{\alpha^2}{2}\int_0^T e^{-2\gamma t}|\dot{\boldsymbol{\theta}}(t)|^2_{\mathbb{R}^N}\,dt\right).$$

Moreover, for any $f\in C^1_{\mathrm{b}}(M;\mathbb{R})$,

$$\text{(6.43)}\qquad (D^M_{\boldsymbol{\theta}(T)})_x \mathbf{P}^M_T f - (D^M_{\boldsymbol{\theta}(0)})_x \mathbf{P}^M_T f = \mathbb{E}^{\mu_{\mathbb{R}^N}}\Big[B_{\boldsymbol{\theta}}(T,x,\mathbf{w})f\big(p(T,x,\mathbf{w})\big)\Big].$$

PROOF: The basic observation on which both parts of this theorem turn is that, for any smooth $F : M \times \mathbb{R} \longrightarrow \mathbb{R}$,

$$\tag{6.44} \begin{aligned} &\frac{1}{2}\sum_{i=1}^{N}\left(\mathfrak{Y}_i^{\epsilon,M}(t)\right)_{(y,O,J,B)} \circ \mathfrak{Y}_i^{\epsilon,M}(t)F \\ &= \left[\tfrac{1}{2}\Delta_M F(\,\cdot\,,B)\right](y) + \frac{\partial D^{\epsilon,M}_{OJ\dot{\boldsymbol{\theta}}(t)}F}{\partial B}(y,B) + \frac{\left|J\dot{\boldsymbol{\theta}}(t)\right|^2_{\mathbb{R}^N}}{2}\frac{\partial^2 F}{\partial B^2}(y,B). \end{aligned}$$

Indeed, the only aspect of the preceding which may cause any difficulty is the absence of the term

$$\left(\sum_{i=1}^{N}\mathfrak{D}_i^{\epsilon,M}\left(OJ\dot{\boldsymbol{\theta}}(t),\Pi^M\mathbf{e}_i\right)_{\mathbb{R}^N}\right)\frac{\partial F(y,\,*\,)}{\partial B}.$$

However, (cf. Lemma 5.3)

$$\begin{aligned} \mathfrak{D}_i^{\epsilon,M}\left(OJ\dot{\boldsymbol{\theta}}(t),\Pi^M\mathbf{e}_i\right)_{\mathbb{R}^N} &= \left(\mathcal{A}(\Pi^M\mathbf{e}_i)OJ\dot{\boldsymbol{\theta}}(t),\Pi^M\mathbf{e}_i\right)_{\mathbb{R}^N} \\ &\qquad + \left(OJ\dot{\boldsymbol{\theta}}(t),\partial\Pi^M(D_i^{\epsilon,M})\mathbf{e}_i\right)_{\mathbb{R}^N} \\ &= \left(OJ\dot{\boldsymbol{\theta}}(t),\partial\Pi^M(D_i^{\epsilon,M})(\Pi^M)^{\perp}\mathbf{e}_i\right)_{\mathbb{R}^N}, \end{aligned}$$

and (cf. (5.21))

$$\sum_{i=1}^{N}\partial\Pi^M(D_i^{\epsilon,M})(\Pi^M)^{\perp}\mathbf{e}_i = \Pi^M\left(\sum_{i=1}^{N}\partial\Pi^M(D_i^{\epsilon,M})\mathbf{e}_i\right) = \Pi^M\mathbf{N} = \mathbf{0}.$$

We now prove (6.42) by applying (6.44) with $F(y,B) = e^{\alpha B}$, thereby concluding that, for any $R \in (0,\infty)$,

$$\begin{aligned} &\exp\Big(\alpha B_{\boldsymbol{\theta}}\big(t\wedge\zeta_R(x),x,\mathbf{w}\big)\Big) \\ &\quad - \frac{\alpha^2}{2}\int_0^{t\wedge\zeta_R(x,\mathbf{w})}\left|\hat{J}(\tau,x,\mathbf{w})\dot{\boldsymbol{\theta}}(\tau)\right|^2_{\mathbb{R}^N}\exp\Big(\alpha B_{\boldsymbol{\theta}}(\tau,x,\mathbf{w})\Big)\,d\tau \end{aligned}$$

is a $\mu_{\mathbb{R}^N}$-martingale, when

$$\zeta_R(x,\mathbf{w}) = \inf\left\{t \ge 0 :\ \mathrm{dist}^M\big(p(t,x,\mathbf{w}),x\big) \vee |B(t,x,\mathbf{w})| \ge R\right\}.$$

In particular, since (cf. (6.20) and the discussion preceding this theorem)

$$\left|\hat{J}(\tau,x,\mathbf{w})\dot{\boldsymbol{\theta}}(\tau)\right|_{\mathbb{R}^N} \le e^{-\gamma\tau}|\dot{\boldsymbol{\theta}}(\tau)|_{\mathbb{R}^N},$$

this proves that

$$\begin{aligned}&\mathbb{E}^{\mu_{\mathbb{R}^N}}\Big[\exp\Big(\alpha B_{\boldsymbol{\theta}}(T\wedge\zeta_R(x),x,\mathbf{w})\Big)\Big]\\&\quad\le\frac{\alpha^2}{2}\int_0^T e^{-2\gamma t}|\dot{\boldsymbol{\theta}}(t)|^2_{\mathbb{R}^N}\,\mathbb{E}^{\mu_{\mathbb{R}^N}}\Big[\exp\Big(\alpha B_{\boldsymbol{\theta}}(t\wedge\zeta_R(x),x,\mathbf{w})\Big)\Big]\,dt.\end{aligned}$$

Hence, (6.42) follows after an application of Gronwall's inequality and Fatou's Lemma.

To prove (6.43), we take $f\in C_c^\infty(M;\mathbb{R})$ and set $f_t=\mathbf{P}^M_{T-t}f$. Then by taking $F(t,y,B)=Bf_t$ and applying (6.44), we find that

$$\begin{aligned}&B_{\boldsymbol{\theta}}\big(t\wedge\zeta_R(x),x,\mathbf{w}\big)f\big(p(t\wedge\zeta_R(x),x,\mathbf{w})\big)\\&\quad-\int_0^{t\wedge\zeta_R(x)}\Big(O(\tau,x,\mathbf{w})\hat{J}(\tau,x,\mathbf{w})\dot{\boldsymbol{\theta}}(\tau),\operatorname{grad}^M_{p(\tau,x,\mathbf{w})}f_t\Big)_{\mathbb{R}^N}\,d\tau\end{aligned}$$

is a $\mu_{\mathbb{R}^N}$-martingale. Morover, after a standard mollification arguement, it becomes clear that the preceding continues to hold for all $f\in C^1_b(M;\mathbb{R})$. In particular, by (6.42) and Lebesgue's Dominated Convergence Theorem, one arrives at

$$\begin{aligned}&\mathbb{E}^{\mu_{\mathbb{R}^N}}\Big[B_{\boldsymbol{\theta}}(T,x,\mathbf{w})f\big(p(T,x,\mathbf{w})\big)\Big]\\&\quad=\mathbb{E}^{\mu_{\mathbb{R}^N}}\left[\int_0^T\Big(O(t,x,\mathbf{w})\hat{J}(t,x,\mathbf{w})\dot{\boldsymbol{\theta}}(t),\operatorname{grad}^M_{p(t,x,\mathbf{w})}f_t\Big)_{\mathbb{R}^N}\,dt\right].\end{aligned}$$

Finally, after taking into account the discussion preceding this theorem and using the second part of Lemma 6.37, we get (6.43). □

Because it is a variation on one which appeared originally in equation (2.70) of the book [3] by J.–M. Bismut, we call (6.43) *Bismut's formula* and the quantity $B_{\boldsymbol{\theta}}(T,x,\mathbf{w})$ the *Bismut factor*. As our derivation makes clear, Bismut's formula is really nothing but another stochastic version of Bochner's identity (6.19). Nonetheless, because f on the right hand side carries no derivatives, it is a particularly interesting and potentially useful version.

6.3.3. Measurability Again. [10] For the reader who has been keeping track of what is and what is not measurable with respect the path $p(\,\cdot\,)$, it should be clear that everything which appears in Bismut's formula (6.43) *is* a measurable function of the path. To check this, recall the construction of the measure $\mathbb{R}_x$ and the path $\mathbf{w}$ made in § 4.3.1, and set $\mathbf{p}_n=\big(p_n(\,\cdot\,),O_{p_n}(\,\cdot\,)\big)$ where p_n is given by (4.36). Then, we will be done when we show that

$$\begin{aligned}\sup_{t\in[0,T]}\bigg|&\int_0^t\Big(O_{p_n}(\tau)\hat{J}_{\mathbf{p}_n}(\tau)\dot{\boldsymbol{\theta}}(\tau),\dot{p}_n(\tau)\Big)_{\mathbb{R}^N}\,d\tau\\&-\int_0^t\Big(O_{p(\,\cdot\,,x,\mathbf{w}_n)}(\tau)\hat{J}_{\mathbf{p}(\,\cdot\,,x,\mathbf{w}_n)}(\tau)\dot{\boldsymbol{\theta}}(\tau),\dot{\mathbf{w}}_n\Big)_{\mathbb{R}^N}\,d\tau\bigg|=0\quad(\text{a.s., }\mathbb{R}_x).\end{aligned}$$

[10] The reader who heeded my advice to skip § 4.3.1 and § 5.1.3 should again take my advice and skip this one.

But, because $O_{p(\,\cdot\,,x,\mathbf{w}_n)}(\tau)\hat{J}_{\mathfrak{p}(\,\cdot\,,x,\mathbf{w}_n)}(\tau)\dot{\boldsymbol{\theta}}(\tau) \in T_{p(\tau,x,\mathbf{w}_n)}M$, the $\dot{\mathbf{w}}_n$ in the second integral can be replaced by $\dot{p}(\tau, x, \mathbf{w}_n)$, and so the desired conclusion can be derived from a repetition of the reasoning given in § 4.3.1 and § 5.1.3.

In view of the preceding, it is reasonable, albeit somewhat irresponsible, to write (cf. (5.9) and (6.14)) $\hat{J}_p = \hat{J}_{\mathfrak{p}}$ where $\mathfrak{p} = (p, O_p)$, and then write the Bismut factor as

$$B_{\boldsymbol{\theta}}(t,p) = \int_0^T \Big(O_p(t)\hat{J}_p(t)\dot{\boldsymbol{\theta}}(t), \dot{p}(t)\Big)_{\mathbb{R}^N}\, dt.$$

After doing this, one discovers an intriguing property about the holonomy along Brownian paths. Namely, from (6.43), it is easy to see that

$$\boldsymbol{\theta}(T) = \boldsymbol{\theta}(0) \implies \mathbb{E}^{\mu_{\mathbb{R}^N}}\Big[B_{\boldsymbol{\theta}}(T,x,\mathbf{w})\,\Big|\,p(T,x,\mathbf{w})\Big] = 0 \quad (\text{a.s., } \mu_{\mathbb{R}^N}).$$

Thus, if we adopt the above expression for the Bismut factor, then we find that

$$\boldsymbol{\theta}(T) = \boldsymbol{\theta}(0) \implies \mathbb{E}^{\mathbb{P}_x^M}\left[\int_0^T \Big(O_p(t)\hat{J}_{\mathfrak{p}}(t)\dot{\boldsymbol{\theta}}(t), \dot{p}(t)\Big)_{\mathbb{R}^N}\, dt\,\Big|\,p(T) = y\right] = 0$$

for all $x, y \in M$. In other words, some sort of miraculous cancelation is taking place in the "random holonomy" along Brownian paths.

6.3.4. An Estimate on Logarithmic Gradients. It should be clear that (6.43) in conjunction with (6.42) leads to estimates on the gradient of $\log \mathbf{P}_T^M f$. In order to extract the best estimate that we can, we will use the following simple application of Jensen's inequality.

6.45 LEMMA. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and f a non-negative, $\mathcal{F}$-measurable function on Ω with expectation value 1. If $\psi : \Omega \longrightarrow \mathbb{R}$ is a $\mathcal{F}$-measurable function such that ψf is $\mathbb{P}$-integrable, then*[11]

$$\mathbb{E}^{\mathbb{P}}[\psi f] \le \mathbb{E}^{\mathbb{P}}\left[f \log f\right] + \log \mathbb{E}^{\mathbb{P}}\left[e^{\psi}\right].$$

PROOF: First note that it suffices to handle the case when f is uniformly positive and both f and ψ are bounded. Thus, make these assumptions, and define $\mathbb{Q}$ to be the probability measure determined by $\frac{d\mathbb{Q}}{d\mathbb{P}} = f$. Then, by Jensen's inequality,

$$\log\Big(\mathbb{E}^{\mathbb{P}}\left[e^{\psi}\right]\Big) = \log\Big(\mathbb{E}^{\mathbb{Q}}\left[f^{-1}e^{\psi}\right]\Big) \le -\mathbb{E}^{\mathbb{Q}}\left[\log f\right] + \mathbb{E}^{\mathbb{Q}}[\psi]. \quad \square$$

[11] Recall that $0 \log 0 = 0$ and note that $\xi \log \xi \ge -e^{-1}$ for all $\xi \in [0, \infty)$.

6.46 THEOREM. *Set*

$$\beta(s) = \frac{s}{e^s - 1} \big(\equiv 1 \text{ when } s = 0\big). \tag{6.47}$$

Then, for each $f \in C_b\big(M;(0,\infty)\big)$,

$$\begin{aligned}\big\|\mathrm{grad}_x^M \log \mathbf{P}_T^M f\big\|^2 &\le \frac{2\beta(2\gamma T)^2}{T}\left[\mathbf{P}_T^M\left(\frac{f}{[\mathbf{P}_T^M f](x)}\log\frac{f}{[\mathbf{P}_T^M f](x)}\right)(x)\right] \\ &\le \frac{2\beta(2\gamma T)^2}{T}\log\frac{\|f\|_{C_b(M;\mathbb{R})}}{[\mathbf{P}_T^M f](x)}.\end{aligned}$$

PROOF: Let $\boldsymbol{\eta} \in \mathbb{R}^N \setminus \{\mathbf{0}\}$ be given, and choose a $\boldsymbol{\theta} \in C^2\big([0,\infty);\mathbb{R}^N\big)$ so that $\boldsymbol{\theta}(0) = \mathbf{0}$ and $\boldsymbol{\theta}(T) = \boldsymbol{\eta}$. Then, by (6.43),

$$\big[D_{\boldsymbol{\eta}}^M \log \mathbf{P}_T^M f\big](x) = \mathbb{E}^{\mu_{\mathbb{R}^N}}\Big[B_{\boldsymbol{\theta}}(T,x,\mathbf{w})\bar{f}_{(T,x)}\big(p(T,x,\mathbf{w})\big)\Big],$$

where $\bar{f}_{(T,x)} \equiv \frac{f}{[\mathbf{P}_T^M f](x)}$. Hence, if $H(T,x) \equiv \mathbb{E}^{\mu_{\mathbb{R}^N}}\big[\bar{f}_{(T,x)}\log\bar{f}_{(T,x)}\big]$, then, by Lemma 6.45 and (6.42), for any $\alpha \in \mathbb{R}$,

$$\begin{aligned}\alpha\big[D_{\boldsymbol{\eta}}^M \log \mathbf{P}_T^M f\big](x) &\le H(T,x) + \log\mathbb{E}^{\mu_{\mathbb{R}^N}}\Big[e^{\alpha B_{\boldsymbol{\theta}}(T,x,\mathbf{w})}\Big] \\ &\le H(T,x) + \frac{\alpha^2}{2}\int_0^T e^{-2\gamma t}\big|\dot{\boldsymbol{\theta}}(t)\big|^2\,dt,\end{aligned}$$

which means that

$$\Big|\big[D_{\boldsymbol{\eta}}^M \log \mathbf{P}_T^M f\big](x)\Big| \le \frac{H(T,x)}{\alpha} + \frac{\alpha}{2}\int_0^T e^{-2\gamma t}\big|\dot{\boldsymbol{\theta}}(t)\big|^2\,dt$$

for all $\alpha > 0$. To get the best estimate from this, we first take

$$\boldsymbol{\theta}(t) = \frac{e^{2\gamma t} - 1}{e^{2\gamma T} - 1}\boldsymbol{\eta}\left(\equiv \frac{t}{T}\boldsymbol{\eta} \text{ when } \gamma = 0\right),$$

since this is the $\boldsymbol{\theta}$ which minimizes the integral subject to the boundary conditions. We finally minimize with respect to $\alpha > 0$ and thereby arrive at the desired estimate. □

It is interesting to compare the estimate obtained in Theorem 6.46 to the one of Li and Yau in (6.27). Clearly, they are both estimates on the logarithmic gradient of the heat flow. Unfortunately, it is not at all clear what else they have in common. In particular, I see no way of passing to theirs from the one here. (See § 10.6)

CHAPTER 7

Some Intrinsic Riemannian Geometry

In the preceding two chapters we developed our theory when M is an embedded submanifold of $\mathbb{R}^N$ for some N. In that submanifolds are the only ones which most mortals can picture, there is a lot to be said for working in that context. Further, there are theorems which say that there is really no loss in generality to restrict ones attention to submanifolds of $\mathbb{R}^N$. For example, there is a relatively simple theorem of H. Whitney [44] which says that any separable, differentiable manifold M can be realized as a submanifold of $\mathbb{R}^N$ for some N. If one wants to work harder, one can show that N can be taken less than or equal to $2\dim(M)$. Further, a much harder theorem, due to J. Nash [31], says that a separable Riemannian manifold of dimension d can be isometrically embedded in $\mathbb{R}^N$ with $N \leq \frac{1}{2}d(d+1)(3d+11)$. On the other hand, experience indicates that there are benefits to be gained from forcing oneself to work intrinsically. In particular, an intrinsic approach often reveals structure which is masked when one relies too heavily on extrinsic considerations. Thus, in this and the following chapters, we will give an intrinsic treatment of the topics in Chapters 4, 5, and 6.

Throughout, M will denote a separable, d-dimensional, C^∞, connected differentiable manifold. Given a separable Hilbert space E, we will use $C^\infty(M;E)$ to stand for the space of infinitely differentiable $f : M \longrightarrow E$. Further, for each $x \in M$, the *tangent space* T_xM *to* M *at* x will be identified with the vector space of linear functionals $X_x : C^\infty(M;\mathbb{R}) \longrightarrow \mathbb{R}$ which satisfy Leibnitz's rule: $X_x(fg) = fX_xg + gX_xf$. In particular, if $\gamma : (a,b) \longrightarrow M$ is a smooth curve, then, for each $t \in (a,b)$, we use $\dot{\gamma}(t)$ to denote the element of $T_{\gamma(t)}M$ such that

$$\dot{\gamma}(t)f = (\gamma_*\partial_t)f \equiv \frac{d}{dt} f \circ \gamma(t)$$

and recall that

$$T_xM = \big\{\dot{\gamma}(0) : \gamma \text{ is a smooth curve from a neighborhood of } 0 \in \mathbb{R} \text{ into } M \text{ with } \gamma(0) = x\big\}.$$

In fact, if (U, Φ) is a coordinate chart at x (i.e., U is an open neighborhood of x and Φ is a diffeomorphism from U to an open subset of $\mathbb{R}^d$), then

$$T_xM = \left\{(\partial_k^\Phi)_{\boldsymbol{\xi}} : \boldsymbol{\xi} \in \mathbb{R}^d\right\},$$

where

$$(7.1) \qquad (\partial_\xi^\Phi)_x f \equiv (\Phi^{-1})_*(\partial_\xi)_{\Phi(x)} f = (\partial_\xi)_{\Phi(x)} f \circ \Phi^{-1},$$

and ∂_ξ is the partial derivative operator on $\mathbb{R}^d$ in the direction ξ. Finally, a vector field X on M is a map $x \in M \longmapsto X_x \in T_x M$ with the property that $x \in M \longmapsto X_x f \in \mathbb{R}$ is smooth (i.e., C^∞) for each $f \in C^\infty(M;\mathbb{R})$. Equivalently, for each coordinate chart (U,Φ), $X_x = \sum_{k=1}^d a_k(x)(\partial_k^\Phi)_x$, where $\{a_k\}_1^d \subseteq C^\infty(U;\mathbb{R})$. In particular, we will often identify a vector field X with the first order, linear differential operator $X : C^\infty(M;\mathbb{R}) \longrightarrow C^\infty(M;\mathbb{R})$ given by $[Xf](x) = X_x f$ at each $x \in M$.

7.1 Diffusions on an Abstract Manifold

Assume that $(X_0,\dots,X_d)$ are smooth vector fields on M with the property that each $X \in \operatorname{span}(X_0,\dots,X_d)$ is complete (in the sense that, for each $x \in M$, there is an integral curve of X which passes through x at time 0 and lives for all time). Next, given $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$, define the polygonal paths $\mathbf{w}_n$, $n \in \mathbb{N}$, as in (3.17), and let $p(\,\cdot\,,x,\mathbf{w}_n)$ be the solution to

$$\dot p(t,x,\mathbf{w}_n) = (X_0)_{p(t,x,\mathbf{w}_n)} + \frac{1}{2}\sum_{k=1}^d \dot w_n^k(t)(X_k)_{p(t,x,\mathbf{w}_n)} \quad \text{with } p(0,x,\mathbf{w}_n) = x.$$

7.1.1. Basic Existence Statement. In spite of the fact that we are attempting to get away from the idea that M has to be a submanifold of $\mathbb{R}^N$, the basic facts about diffusions on an abstract differentiable manifold M are most easily proved by invoking the familiar theorem of H. Whitney alluded to earlier, the one which says that M can always be realized as a submanifold of $\mathbb{R}^{2d}$. To be more precise, his theorem says that if M is a separable, d-dimensional manifold, then M is diffeomorphic to a closed, embedded submanifold of $\mathbb{R}^{2d}$. Thus there is no problem about repeating the reasoning in § 4.1.1 to prove the following statement about the relationship between the martingale problem for the operator $\mathcal{L}$ determined (cf. (2.3)) by $(X_0,\dots,X_d)$ and the $\mu_{\mathbb{R}^d}$-almost sure existence of limits of $\{p(\,\cdot\,,x,\mathbf{w}_n)\}_1^\infty$.

7.2 THEOREM. *For each $x \in M$, there is at most one solution $\mathbb{P}_x^{\mathcal{L}} \in \mathbf{M}_1(\mathcal{P}(M))$ to the martingale problem for $\mathcal{L}$ starting at x. Moreover, the martingale problem for $\mathcal{L}$ at x is well-posed if and only if there exists a measurable $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto p(\,\cdot\,,x,\mathbf{w}) \in \mathcal{P}(M)$ such that*

$$p(\,\cdot\,,x,\mathbf{w}_n) \longrightarrow p(\,\cdot\,,x,\mathbf{w}) \quad \text{in } \mathcal{P}(M) \text{ for } \mu_{\mathbb{R}^d}\text{-almost every } \mathbf{w},$$

in which case $\mathbb{P}_x^{\mathcal{L}}$ is the $\mu_{\mathbb{R}^d}$-distribution of $\mathbf{w} \rightsquigarrow p(\,\cdot\,,x,\mathbf{w})$. In fact, if the martingale problem for $\mathcal{L}$ is well-posed on M, then the map $\mathbf{w} \rightsquigarrow p(\,\cdot\,,*\,,\mathbf{w})$ can be chosen so that

(1) *for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$*

$$p(\,\cdot\,,*\,,\mathbf{w}_n) \longrightarrow p(\,\cdot\,,*\,,\mathbf{w}) \text{ in } C^{0,\infty}\big([0,\infty)\times M; M\big),$$

(2) *for each* $x \in M$, *the* $\mu_{\mathbb{R}^d}$*-distribution of* $\mathbf{w} \rightsquigarrow p(\cdot, x, \mathbf{w}) \in \mathcal{P}(M)$ *is* $\mathbb{P}_x^{\mathcal{L}}$,
(3) *for* $\mu_{\mathbb{R}^d}$*-almost every* $\mathbf{w}$, $p(\cdot, *, \mathbf{w}) \in C^{0,\infty}([0,\infty) \times M; M)$ *and, for all* $t \in [0,\infty)$, $p(t, *, \mathbf{w})$ *is one-to-one and* $p(t, *, \mathbf{w})_*$ *is nowhere degenerate.*

In particular, if the martingale problem for $\mathcal{L}$ *is well-posed on* M, *then* $x \in M \longmapsto \mathbb{P}_x^{\mathcal{L}} \in \mathbf{M}_1(\mathcal{P}(M))$ *is weakly continuous and (3.16) holds for every* $\{\mathcal{F}_t : t \geq 0\}$ *stopping time* $\zeta : \mathcal{P}(M) \longrightarrow [0,\infty]$. *Finally, if, in addition,* $\check{\mathcal{L}}$ *is given by (3.45) and the martingale problem for* $\check{\mathcal{L}}$ *is also well-posed on* M, *then, for each* $t \in [0,\infty)$, $p(t, *, \mathbf{w}) \in \mathrm{Diff}(M; M)$ *for* $\mu_{\mathbb{R}^d}$*-almost every* $\mathbf{w}$.

7.2 Riemannian Manifolds

A Riemannian metric on a differentiable manifold M is a smoothly varying assignment of an inner product to the tangent space at each point $x \in M$. That is, for each $x \in M$, we have an inner product $(X_x, Y_x) \in T_xM \times T_xM \longmapsto \langle X_x, Y_x\rangle \in \mathbb{R}$ on T_xM which varies smoothly in the sense that, for all smooth vector fields X and Y on M, $x \rightsquigarrow \langle X_x, Y_x\rangle$ is a smooth function. A manifold M with a specified Riemann metric $x \rightsquigarrow \langle\cdot\,,\cdot\rangle$ is called a *Riemann manifold*, and the quantity

$$\|X_x\| \equiv \sqrt{\langle X_x, X_x\rangle}$$

is called the *length* of X_x. Since, by Whitney's Theorem, we may always realize M as a submanifold of some $\mathbb{R}^N$, there is no obstruction to equipping a manifold with a Riemannian metric. The challenge comes when one wants to have the Riemannian metric possess prescribed properties.

7.2.1. Basic Quantities. For most computations, it is necessary to provide a local description of the Riemann metric. Thus, given a coordinate chart (U, Φ), define the smooth map $x \in U \longmapsto g^\Phi(x) \in \mathrm{Hom}(\mathbb{R}^d; \mathbb{R}^d)$ by (cf. (7.1))

$$(7.3) \qquad \left(g^\Phi(x)\boldsymbol{\xi}, \boldsymbol{\xi}'\right)_{\mathbb{R}^d} = \left\langle (\partial_{\boldsymbol{\xi}}^\Phi)_x, (\partial_{\boldsymbol{\xi}'}^\Phi)_x \right\rangle, \quad x \in U \text{ and } \boldsymbol{\xi},\, \boldsymbol{\xi}' \in \mathbb{R}^d.$$

Clearly, for every $x \in U$, $g^\Phi(x)$ is symmetric and (strictly) positive definite.

Once M has a Riemannian metric, we have everything that is needed to reproduce the results in § 4.1.2. Namely, given a coordinate chart (U, Φ) and an (Euclidean) orthonormal basis $\mathfrak{e} = (\mathbf{e}_1, \ldots, \mathbf{e}_d)$ in $\mathbb{R}^d$, we set

$$(7.4) \qquad \det\left(g^\Phi(x)\right) = \det\left(\left(\left(g^\Phi(x)\mathbf{e}_i, \mathbf{e}_j\right)_{\mathbb{R}^d}\right)\right)_{1 \leq i,j \leq d},$$

which is justified by the observation that the right hand side is independent of the choice of orthonormal basis $\mathfrak{e}$. Next, we define a measure λ_M, called the *Riemann measure*, by the prescription given in (4.5), in which case the Jacobi change of variable formula in Theorem 4.11 applies. At the same time, if we use the prescription in (4.13) and therefore define the divergence $\mathrm{div}^M(X)$ of a vector field X by

$$(7.5) \qquad \mathrm{div}^M(X) = \frac{1}{\sqrt{\det(g^\Phi)}} \sum_{i,j=1}^d \partial_i^\Phi\left(\sqrt{\det(g^\Phi)}\,\left(g^\Phi\right)^{i,j}\langle X, \partial_j^\Phi\rangle\right)$$

(where the use of superscripts on $(g^\Phi)^{i,j}$ indicates that we are dealing with the inverse matrix $(g^\Phi)^{-1}$), then (4.14), (4.15) as well as Theorem 4.17, Corollary 4.19, and Theorem 4.21 all follow by the same arguments as we used there.

Finally, given $f \in C^\infty(M;\mathbb{R})$, we define first the gradient $\mathrm{grad}^M f$ to be the vector field determined by (cf. (4.23))

$$(X_x, \mathrm{grad}_x^M f) = X_x f \quad \text{for } x \in M \text{ and } X_x \in T_xM \tag{7.6}$$

and then take the *Laplacian* $\Delta_M f$ to be given by the prescription in (4.22). In particular, (4.25) holds without any change, and (4.26) gets replaced here by

$$-\int_M f\Delta_M g\, d\lambda_M = \int_M (\mathrm{grad}^M f, \mathrm{grad}^M g)\, d\lambda_M = -\int_M g\Delta_M f\, d\lambda_M \tag{7.7}$$

for $f, g \in C_c^\infty(M;\mathbb{R})$.

7.2.2. The Levi–Civita Connection. In view of the preceding, it is clear that our next order of business is the introduction of an appropriate notion of parallel transport, and for this purpose we will reverse the order taken in Chapter 5 and adopt the conventional modern approach which describes parallel transport in terms of a Riemannian connection. To this end, we will say that ∇^M is the *Levi-Civita Riemannian connection* on M if $(X,Y) \rightsquigarrow \nabla_X^M Y$ takes pairs of vector fields on M into vector fields on M in such a way that, for each $x \in M$,

$$X_x \in T_xM \longmapsto \nabla_{X_x}^M Y \in T_xM \quad \text{is linear} \tag{7.8}$$

$$\begin{aligned} &\nabla_{X_x}^M (Y+Z) = \nabla_{X_x}^M Y + \nabla_{X_x}^M Z \\ &\nabla_{X_x}^M (fY) = (X_x f)Y_x + f(x)\nabla_{X_x}^M Y \end{aligned} \tag{7.9}$$

$$X_x(Y,Z) = (\nabla_{X_x}^M Y, Z_x) + (Y_x, \nabla_{X_x}^M Y) \tag{7.10}$$

$$\nabla_X^M Y - \nabla_Y^M X = [X,Y]. \tag{7.11}$$

The algebraic properties (7.8) and (7.9) are common to all connections, (7.10) says that the connection is Riemannian in that it is adapted to the given Riemann metric. and (7.11) is a technical condition which says that the Levi-Civita connection is *torsion free*, the property which distinguishes it form all other Riemannian connections. The existence and uniqueness of ∇^M are basic and familiar facts of modern Riemannian geometry. For example, suppose that $x \rightsquigarrow \mathbf{e}(x) = \big((E_1)_x, \dots, (E_d)_x\big)$ is a smoothly varying choice of orthonormal bases for points in an open set U[1]. If ∇^M exists, then, by (7.10), $(\nabla_{E_i}^M E_j, E_k) =$

[1] For example, one can start with a coordinate chart (U, Φ) and take $\mathbf{e}(x)$ to be the orthonormal basis obtained by applying Gram-Schmidt to $\big((\partial_{\mathbf{e}_1}^\Phi)_x, \dots, (\partial_{\mathbf{e}_d}^\Phi)_x\big)$ at each $x \in U$ for some orthonormal basis $(\mathbf{e}_1, \dots, \mathbf{e}_d)$ in $\mathbb{R}^d$.

$-(\nabla^M_{E_i} E_k, E_j)$, and so, by (7.11),

$$\begin{aligned}
(\nabla^M_{E_i} E_j, E_k) &= \langle [E_i, E_j], E_k \rangle + (\nabla^M_{E_j} E_i, E_k) \\
&= \langle [E_i, E_j], E_k \rangle - (\nabla^M_{E_j} E_k, E_i) \\
&= \langle [E_i, E_j], E_k \rangle - \langle [E_j, E_k], E_i \rangle - (\nabla^M_{E_k} E_j, E_i) \\
&= \langle [E_i, E_j], E_k \rangle - \langle [E_j, E_k], E_i \rangle + \langle [E_k, E_i], E_j \rangle + (\nabla^M_{E_i} E_k, E_j) \\
&= \langle [E_i, E_j], E_k \rangle - \langle [E_j, E_k], E_i \rangle + \langle [E_k, E_i], E_j \rangle - (\nabla^M_{E_i} E_j, E_k).
\end{aligned}$$

Hence, if it exists, then

$$\begin{aligned}
(\nabla^M_{E_i} E_j, E_k) &= \Gamma^k_{i,j}(\mathbf{e}) \\
&\equiv \frac{1}{2}\Big(\langle [E_i, E_j], E_k \rangle - \langle [E_j, E_k], E_i \rangle + \langle [E_k, E_i], E_j \rangle \Big).
\end{aligned} \tag{7.12}$$

Obviously, since every point in M admits an open neighborhood U on which such an $x \rightsquigarrow \mathbf{e}(x)$ exists, this proves uniqueness. In addition, if one starts with (7.12) and uses (7.8) and (7.9) to extend the definition to all vector fields on U, it is an easy, if somewhat tedious, exercise to see that the extension satisfies (7.10) and (7.11). That is, ∇^M both exists and is unique.

In keeping with the terminology introduced in § 5.2, we call $\nabla^M_{X_x} Y$ the *covariant derivative* of Y with respect of X_x. The quantities $\Gamma^k_{i,j}(\mathbf{e})$ are called the *Christoffel symbols* of ∇^M relative to $x \rightsquigarrow \mathbf{e}(x)$. Notice that because we took $\mathbf{e}(x)$ to be orthonormal at each $x \in U$, these Christoffel symbols satisfy the skew symmetry property

$$\Gamma^k_{i,j}(\mathbf{e}) = -\Gamma^j_{i,k}(\mathbf{e}). \tag{7.13}$$

7.2.3. Parallel Transport. Now that we have a Riemannian connection, we will produce a notion of parallel transport for which (5.11) holds, and the first step is to extend the definition of covariant differentiation to vector fields which are defined along a smooth curve. Namely, suppose that $p : [a, b] \longrightarrow M$ is a smooth curve. We say that $t \in [a, b] \longmapsto Y(t) \in T_{p(t)}M$ is smooth if $t \in [a, b] \longmapsto Y(t)f \in \mathbb{R}$ is smooth for every $f \in C^\infty(M; \mathbb{R})$. What we want to do is define the covariant derivative $\frac{DY}{dt}$ of Y along p so that

$$\begin{aligned}
\frac{DY_{p(t)}}{dt}(t) &= \nabla^M_{\dot p(t)} Y \quad \text{if } Y \text{ is a vector field on } M \\
\frac{D(\alpha Y)}{dt}(t) &= \dot\alpha(t) Y(t) + \alpha(t) \frac{DY}{dt}(t) \quad \text{if } \alpha \in C^\infty\big([a, b]; \mathbb{R}\big).
\end{aligned} \tag{7.14}$$

To see that there is precisely one way to do this, note that the question is completely local. Thus, we may and will assume that p lies entirely in an open neighborhood U on which there is a smoothly varying selection $x \in U \longmapsto \mathbf{e}(x) = \big((E_1)_x, \ldots, (E_d)_x\big) \in (T_xM)^d$ such that $\mathbf{e}(x)$ is an orthonormal basis for each $x \in U$. It is then easy to see that $t \rightsquigarrow Y(t)$ is a smooth vector field along p if and only if

$$t \in [a,b] \longmapsto \alpha^k(t) \equiv \langle Y(t), (E_k)_{p(t)} \rangle \in \mathbb{R}$$

is smooth for each $1 \leq k \leq d$, in which case (7.14) leaves us no alternative to taking

$$\frac{DY}{dt}(t) = \sum_{k=1}^{d} \Big(\dot{\alpha}^k(t)(E_k)_{p(t)} + \alpha^k(t) \nabla^M_{\dot{p}(t)} E_k \Big).$$

Notice that the preceding expression also shows that (7.10) transfers to this setting as the Leibnitz rule

$$\frac{d}{dt}\langle Y(t), Z(t) \rangle = \left\langle \frac{DY}{dt}(t), Z(t) \right\rangle + \left\langle Y(t), \frac{DZ}{dt}(t) \right\rangle \tag{7.15}$$

for any pair of smooth vector fields along p.

With the preceding definition at hand, we say that $Y_{p(b)} \in T_{p(b)}M$ is obtained from $Y_{p(a)} \in T_{p(a)}M$ by *parallel transport* along p and write $Y_{p(b)} = \mathcal{T}_p Y_{p(a)}$ if $Y_{p(b)} = Y(b)$ where $t \in [a,b] \longmapsto Y(t) \in T_{p(t)}M$ satisfies

$$\frac{DY}{dt}(t) = 0 \quad \text{for } t \in (a,b) \text{ with } Y(0) = Y_{p(a)}.$$

To see that for each $Y_{p(a)} \in T_{p(a)}M$ there is a unique $Y_{p(b)} = \mathcal{T}_p Y_{p(a)}$, we may and will assume that $t \in [a,b] \longmapsto p(t) \in U$, where U is again an open set on which there exists a smooth choice of $x \rightsquigarrow \mathbf{e}(x)$. But then the problem comes down to showing that there is exactly one solution to (cf. (7.12))

$$\frac{d\alpha^k}{dt}(t) + \sum_{i,j=1}^{d} \langle \dot{p}(t), (E_i)_{p(t)} \rangle \alpha^j(t) \big[\Gamma^k_{i,j}(\mathbf{e})\big](p(t)) = 0$$

with $\alpha^k(a) = \langle Y_a, (E_k)_{p(a)} \rangle$ for $1 \leq k \leq d$, in which case $\mathcal{T}_p Y_{p(a)} = Y(b)$ where

$$t \in [a,b] \longmapsto Y(t) = \sum_{k=1}^{d} \alpha^k(t)(E_k)_{p(t)} \in T_{p(t)}M.$$

Finally, both the existence and uniqueness of $t \in [a,b] \longmapsto \big(\alpha^1(t), \dots, \alpha^d(t)\big) \in \mathbb{R}^d$ are elementary applications of the theory of linear ordinary differential equations. Hence, we have now shown that $\mathcal{T}_p : T_{p(a)}M \longrightarrow T_{p(b)}M$ is well-defined. We summarize a few elementary properties of $\mathcal{T}_p$ in the following.

7.16 LEMMA. *Given $p \in C^\infty([a,b]; M)$, $\mathcal{T}_p$ is a linear isometry from $T_{p(a)}M$ onto $T_{p(b)}M$. Moreover, if $a < c < b$, then we have the composition rule*

$$\mathcal{T}_p = \mathcal{T}_{p\restriction[c,b]} \circ \mathcal{T}_{p\restriction[a,c]}.$$

Finally, for any smooth vector field $t \rightsquigarrow Y(t)$ along p, (5.16) holds. In particular, ∇^M satisfies (5.11).

PROOF: The linearity is a trivial consequence of the linearity of the equations involved. At the same time, the composition property follows from uniqueness: if $Y(t) = \mathcal{T}_{p\upharpoonright[a,t]}Y_{p(a)}$ for $t \in [a,b]$, then $Y(c) = \mathcal{T}_{p\upharpoonright[a,c]}Y_{p(a)}$ and $\frac{DY}{dt}(t) = \mathbf{0}$ for $t \in [c,b]$, and therefore $Y(b) = \mathcal{T}_{p\upharpoonright[c,b]} \circ \mathcal{T}_{p\upharpoonright[a,c]}Y_{p(a)}$. To prove that $\mathcal{T}_p$ is an isometry, we use (7.15) to write

$$\frac{d}{dt}\langle Y(t), Y(t)\rangle = 2\left\langle \frac{DY}{dt}(t), Y(t)\right\rangle = 0,$$

where, again, $Y(t) = \mathcal{T}_{p\upharpoonright[a,t]}Y_{p(a)}$.

Finally, in proving (5.16), first observe that, by the composition rule, it suffices to handle $t = a$. Next, set $x = p(a)$, choose an orthonormal basis $\big((E_1)_x, \dots, (E_d)_x\big)$ in T_xM, and set $E_k(t) = \mathcal{T}_{p\upharpoonright[a,t]}(E_k)_x$. Then, by the isometry property and (7.15),

$$\frac{d}{dt}\langle \mathcal{T}^{-1}_{p\upharpoonright[a,t]}Y(t), (E_k)_x\rangle\Big|_{t=0} = \frac{d}{dt}\langle Y(t), E_k(t)\rangle\Big|_{t=0} = \left\langle \frac{DY}{dt}(0), (E_k)_x\right\rangle,$$

since $\frac{DE_k}{dt}(t) = \mathbf{0}$. $\square$

It may be worth mentioning the significance of (7.13) here. Namely, assume that p lies entirely in an open set U on which there exists a smoothly varying choice $x \rightsquigarrow \mathbf{e}(x) = \big((E_1)_x, \dots, (E_d)_x\big)$ of orthonormal bases, and determine $t \in [a,b] \longmapsto O_p(t) = \big((O_p(t)^k_j)\big) \in \mathrm{Hom}(\mathbb{R}^d;\mathbb{R}^d)$ by $O_p(a) = I$ and

$$\text{(7.17)} \quad \dot{O}_p(t)^k_j + \sum_{i,\ell=1}^d \langle \dot{p}(t), (E_i)_{p(t)}\rangle \big[\Gamma^k_{i,\ell}(\mathbf{e})\big]\big(p(t)\big) O_p(t)^\ell_j = 0, \quad 1 \le j,k \le d,$$

for $t \in (a,b]$. Then, because of the skew symmetry property in (7.13), $O_p(t)$ is orthogonal for all $t \in [a,b]$. Moreover, given $X_{p(a)} \in T_{p(a)}$,

$$\text{(7.18)} \qquad \mathcal{T}_p X_{p(a)} = \sum_{j,k=1}^d O_p(b)^k_j \langle X_{p(a)}, (E_j)_{p(a)}\rangle (E_k)_{p(b)}.$$

To see this, all that one has to do is consider the vector field $t \rightsquigarrow X(t)$ along p which is obtained by taking $X(t)$ equal to the right hand side of (7.18) with b replaced by t and then check that $\frac{DX}{dt}(t) \equiv \mathbf{0}$. What (7.18) is saying is that, although $t \rightsquigarrow (E_j)_{p(t)}$ is not being parallel transported along p, $t \rightsquigarrow O_p(t)$ provides just the right orthogonal transformations so that $t \rightsquigarrow \sum_{k=1}^d O_p(t)^k_j (E_k)_{p(t)}$ is.

7.2.4. An Alternative Expression for the Divergence. Our goal here is to prove that, for any vector field X, $x \in M$, and orthonormal basis $\big((E_1)_x, \dots, (E_d)_x\big)$ in T_xM,

$$\text{(7.19)} \qquad \mathrm{div}^M_x(X) = \sum_{k=1}^d \langle (E_k)_x, \nabla^M_{(E_k)_x} X\rangle.$$

To prove (7.19), first note that the right hand side does not depend on the choice of orthonormal basis. In particular, we can make the choice depend on the x under consideration. With this in mind, choose an orthonormal basis $(\mathbf{e}_1, \dots, \mathbf{e}_d)$ in $\mathbb{R}^d$, a coordinate chart (U, Φ) with $U \ni x$ and $g^\Phi(x) = I$, let σ denote the symmetric square root of g^Φ, and set $E_k = \sum_{j=1}^d \sigma^{k,j}\partial^\Phi_{\mathbf{e}_j}$, where the superscripts mean that we are dealing with the entries of the inverse matrix σ^{-1}. By construction, $\big((E_1)_y, \dots (E_d)_y\big)$ is an orthonormal basis in T_yM at each $y \in U$. In addition, because $g^\Phi(x) = I$, (cf. (7.5))

$$\begin{aligned}\operatorname{div}^M_x(X) &= \sum_{i,j=1}^d (\partial^\Phi_{\mathbf{e}_i})_x\Big(|\sigma|(\sigma^2)^{i,j}\langle X, \partial^\Phi_{\mathbf{e}_j}\rangle\Big) = \sum_{i,k=1}^d (\partial^\Phi_{\mathbf{e}_i} i)_x\Big(|\sigma|(\sigma)^{i,k}\langle X, E_k\rangle\Big)\\ &= \sum_{k=1}^d (E_k)_x\langle X, E_k\rangle + \sum_{k=1}^d \langle X_x, (E_k)_x\rangle\left((E_k)_x|\sigma| + \sum_{i=1}^d (\partial^\Phi_{\mathbf{e}_i})_x\sigma^{i,k}\right),\end{aligned}$$

where $|\sigma| \equiv \det(\sigma) = \sqrt{|g^\Phi|}$. On the other hand,

$$\sum_{k=1}^d \langle (E_k)_x, \nabla^M_{(E_k)_x} X\rangle = \sum_{k=1}^d (E_k)_x\langle X, E_k\rangle - \sum_{i=1}^d \langle X_x, \nabla^M_{(E_i)_x} E_i\rangle,$$

and so, since

$$\begin{aligned}-\sum_{i=1}^d \langle X_x, \nabla^M_{(E_i)_x} E_i\rangle &= -\sum_{k=1}^d \langle X_x, (E_k)_x\rangle \sum_{i=1}^d \langle (E_k)_x, \nabla^M_{(E_i)_x} E_i\rangle\\ &= \sum_{k=1}^d \langle X_x, (E_k)_x\rangle \sum_{i=1}^d \langle \nabla^M_{(E_i)_x} E_k, (E_i)_x\rangle\end{aligned}$$

we are left with proving that

$$(*) \qquad \sum_{i=1}^d \langle \nabla^M_{(E_i)_x} E_k, (E_i)_x\rangle = (E_k)_x|\sigma| + \sum_{i=1}^d (\partial^\Phi_{\mathbf{e}_i})_x\sigma^{i,k}$$

for each $1 \le k \le d$.

To prove $(*)$, we use (7.12) and $\sigma(x) = I$ to see that,

$$\sum_{i=1}^d \langle \nabla^M_{(E_i)_x} E_k, (E_i)_x\rangle = \sum_{i=1}^d \big\langle [E_i, E_k]_x, (E_i)_x\big\rangle = \sum_{i=1}^d (\partial^\Phi_{\mathbf{e}_i})_x\sigma^{i,k} - \sum_{i=1}^d (\partial^\Phi_{\mathbf{e}_k})_x\sigma^{i,i}.$$

Finally, by Cramer's rule, if $\sigma^{(i,j)} = |\sigma|\sigma^{i,j}$ is the (i,j) cofactor of σ, then

$$E_k|\sigma| = \sum_{i,j=1}^d \sigma^{(i,j)}E_k\sigma_{i,j} = |\sigma|\sum_{i,j=1}^d \sigma^{i,j}E_k\sigma_{i,j} = -|\sigma|\sum_{i,j=1}^d \sigma_{i,j}E_k\sigma^{i,j},$$

and so, since $(E_k)_x = (\partial^\Phi_{\mathbf{e}_k})_x$ and $\sigma(x) = I$, we have completed the verification of $(*)$.

7.2.5. The Laplacian as the Trace of the Hessian. As an application of the preceding and in preparation for the construction of Brownian motion on M, we want to develop another representation of the Laplacian Δ_M, this time in terms of the Hessian of a function.

Given $f \in C_c^\infty(M;\mathbb{R})$, $x \in M$, define the *Hessian* of f at x to be the element $H_x^M f \in \mathrm{Hom}(T_xM;T_xM)$ given by

$$H_x^M f X_x = \nabla_{X_x}^M \mathrm{grad}^M f, \quad X_x \in T_xM. \tag{7.20}$$

Although (7.20) makes it explicit that $H_x^M f$ is well-defined as a linear transformation on T_xM, for computational purposes it is important to have another representation of this transformation. Namely, given vector fields X and Y, notice (cf. (7.10)) that

$$(Y, H^M f\, X) = (Y, \nabla_X^M \mathrm{grad}^M f) = X(Y, \mathrm{grad}^M f) - (\nabla_X^M Y, \mathrm{grad}^M f)$$

and conclude that

$$(Y, H^M f\, X) = XYf - \nabla_X^M Y f. \tag{7.21}$$

The expression in (7.21) has the disadvantage that it makes $(Y_x, H^M f_x X_x)$ appear to depend on X and Y in a neighborhood of x, whereas we know from (7.20) that depends only on X_x and Y_x. On the other hand, (7.21) has the advantage that it shows $H_x^M f$ is symmetric in the sense that

$$(Y_x, H_x^M f\, X_x) = (X_x, H_x^M f\, Y_x), \quad X_x, Y_x \in T_xM. \tag{7.22}$$

Indeed, (7.11) is equivalent to $XYf - \nabla_X^M Y f = YXf - \nabla_Y^M X f$.

The immediate reason for our interest in the Hessian is that, in conjuction with (7.19), it gives us another representation of the Laplacian of a function. Namely, given an open set U on which there exists a smooth selection $x \rightsquigarrow \big((E_1)_x, \dots, (E_d)_x\big)$ of orthonormal bases, first observe that $\mathrm{grad}^M f = \sum_{k=1}^d (E_k f)\, E_k$ on U, and then use (4.22), (7.19), and (7.20) to conclude that

$$\Delta_M f \equiv \mathrm{div}^M\big(\mathrm{grad}^M f\big) = \sum_{k,\ell=1}^d \big\langle \nabla_{E_k}^M \mathrm{grad}^M f, E_k \big\rangle,$$

which, by (7.19), means that we now know

$$\Delta_M f = \mathrm{Trace}^M\big(H^M f\big) = \sum_{k=1}^d \big\langle E_k, H^M f\, E_k \big\rangle, \tag{7.23}$$

where, as usual, the trace of a linear transformation on a Hilbert space is computed as the sum of the diagonal elements of the matrix representation with respect to an orthonormal basis. Of course, the first equality in (7.23) gives further evidence that $\Delta_M f$ (cf. (4.25)) is independent of the coordinate system. In addition, one should observe that, since $-(E_k)^{*\lambda_M} = E_k + \mathrm{div}^M(E_k)$ and

$$\mathrm{div}^M(E_k) = \sum_{\ell=1}^{d} \langle \nabla^M_{E_\ell} E_k, E_\ell \rangle = -\sum_{\ell=1}^{d} \langle E_k, \nabla^M_{E_\ell} E_\ell \rangle,$$

the second equality in (7.23) is equivalent to the expression

$$\Delta_M = -\sum_{k=1}^{d} (E_k)^{*\lambda_M} \circ E_k \tag{7.24}$$

for the Laplacian on U. In particular, this expression has the virtue that it brings out the symmetry in $L^2(\lambda_M;\mathbb{R})$ of Δ_M on $C_c^\infty(U;\mathbb{R})$.

7.3 Brownian Motion on M

In this section we will give a "poor man's" construction of Brownian motion on a Riemannian manifold M. The weakness in the construction is that it is entirely local and produces a process from which it is very difficult to extract much geometric information. For this reason we will, in the next chapter, give an entirely different approach, one which is both global and has geometric significance.

7.3.1. Localizing the Laplacian. Let $\{W_n\}_1^\infty$ be a non-decreasing exhaustion of M by relatively compact open sets satisfying $\bar{W}_n \subseteq W_{n+1}$, and choose a countable cover of M by open sets U_α with the properties that

(1) For each n, the set $A_n \equiv \{\alpha : U_\alpha \cap \bar{W}_n \neq \emptyset\}$ is finite.
(2) If $U_\alpha \cap \bar{W}_n \neq \emptyset$, then $\bar{U}_\alpha \subseteq W_{n+1}$.
(3) There exists a smoothly varying selection

$$x \in U_\alpha \longmapsto \mathrm{e}^\alpha(x) = \big((E_1^\alpha)_x, \dots, (E_d^\alpha)_x\big) \in (T_xM)^d$$

of orthonormal bases on U_α.

Next, choose $\eta_\alpha \in C_c^\infty(U_\alpha;\mathbb{R})$ for each α so that $\{\eta_\alpha^2\}$ forms a partition of unity which is subordinate to $\{U_\alpha\}$, and set

$$\rho_n = \sum_{\alpha \in A_n} \eta_\alpha^2 \quad \text{and} \quad \mathcal{L}_n f \equiv \frac{1}{2}\mathrm{div}^M\big(\rho_n \mathrm{grad}^M f\big).$$

Clearly, for each n, $\mathcal{L}_n f = \Delta_M f$ when $f \in C_c^\infty(W_n;\mathbb{R})$. In addition,

$$\mathcal{L}_n = -\frac{1}{2}\sum_{\alpha\in A_n}\sum_{k=1}^{d} (\eta_\alpha E_k^\alpha)^{*\lambda_M} \circ (\eta_\alpha E_k^\alpha). \tag{7.25}$$

To check this, note that $\left(\eta_\alpha E_k^\alpha\right)^{*\lambda_M}\varphi = (E_k^\alpha)^{*\lambda_M}(\eta_\alpha\varphi)$ and therefore (cf. (7.24))

$$
\begin{aligned}
-\sum_{k=1}^d &\left(\eta_\alpha E_k^\alpha\right)^{*\lambda_M} \circ \left(\eta_\alpha E_k^\alpha\right) f \\
&= -\sum_{k=1}^d (E_k^\alpha)^{*\lambda_M}\left(\eta_\alpha^2 E_k^\alpha f\right) = \eta_\alpha^2 \Delta_M f + \left(\operatorname{grad}^M \eta_a^2, \operatorname{grad}^M f\right) \\
&= \eta_\alpha^2 \operatorname{div}^M\left(\operatorname{grad}^M f\right) + \left(\operatorname{grad}^M \eta_a^2, \operatorname{grad}^M f\right) = \operatorname{div}^M\left(\eta_\alpha^2 \operatorname{grad}^M f\right).
\end{aligned}
$$

7.3.2. Construction of Brownian Motion via Localization. With the preceding computations at hand, we can now carry out a construction of Brownian motion[2] by following the kind of procedure developed in § 3.2. Namely, for each n and $x \in W_n$, let $\mathbb{P}_x^n$ be the solution to the martingale problem for $\mathcal{L}_n$ starting at x. Then, because $\mathcal{L}_{n+1} f = \mathcal{L}_n f$ for $f \in C_c^\infty(W_n;\mathbb{R})$, the obvious extension of Theorem 3.10 to the manifold setting tells us that

$$
\mathbb{P}_x^{n+1} \restriction \mathcal{F}_{\zeta^n} = \mathbb{P}_x^n \restriction \mathcal{F}_{\zeta^n}, \quad \text{where } \zeta^n(p) \equiv \inf\{t \geq 0 : p(t) \notin W_n\}.
$$

Moreover,

$$
f\left(p(t\wedge\zeta^n)\right) - \int_0^{t\wedge\zeta^n} \tfrac{1}{2}\Delta_M f\left(p(\tau)\right) d\tau
$$

is a $\mathbb{P}_x^n$-martingale for every $f \in C_c^\infty(M;\mathbb{R})$. Thus, what we would like to do is define $\mathbb{P}_x^M$ on $\mathcal{P}(M)$ by taking it to be the measure which, for n with $W_n \ni x$, coincides with $\mathbb{P}_x^n$ on $\mathcal{F}_{\zeta^n}$. However, this plan has a serious problem when explosion occurs. To overcome this objection, when M is non-compact we introduce the one-point compactification $\hat{M} = M \cup \{\infty\}$ of M, endow $\hat{M}$ with the topology for which $\{\bar{W}_n\complement\}_1^\infty$ is a neighborhood basis at ∞, and set

$$
\hat{\mathcal{P}}(M) = \left\{\hat{p} \in C\left([0,\infty);\hat{M}\right) : \hat{p}(t) = \infty \text{ if } \infty > t \geq \mathfrak{e}(\hat{p}) \equiv \lim_{n\to\infty} \zeta^n(\hat{p})\right\}.
$$

With this definition in place, we can now define a unique Borel probability measure $\mathbb{P}_x^{\hat{M}}$ on $\mathcal{P}(\hat{M})$ by the prescription that

$$
W_n \ni x \implies \mathbb{P}_x^{\hat{M}} \restriction \mathcal{F}_{\zeta^n} = \mathbb{P}_x^n \restriction \mathcal{F}_{\zeta^n} \quad \text{for all } n. \tag{7.26}
$$

We have just proved the first part of the following theorem. The second part follows from (7.25) and the results in § 3.6.2, especially Theorem 3.58.

[2] That is, the construction of solutions $\mathbb{P}_x^M = \mathbb{P}_x^{\frac{1}{2}\Delta_M}$ to the martingale problem for $\frac{1}{2}\Delta_M$.

7.27 THEOREM. *For each $x \in M$, the martingale problem for $\frac{1}{2}\Delta_M$ has at most one solution. Moreover, such a solution exists if and only if*

$$\mathbb{P}_x^n\left(\zeta^n \leq T\right) \searrow 0 \quad \text{for each } T \in [0,\infty),$$

in which case (cf. (7.26)) $\mathbb{P}_x^M = \mathbb{P}_x^{\hat{M}} \restriction \mathcal{P}(M)$. Moreover, if the martingale problem for $\frac{1}{2}\Delta_M$ is well-posed on M, then $x \rightsquigarrow \mathbb{P}_x^M$ is measurable and the associated Markov semigroup $\{\mathbf{P}_t^M : t \geq 0\}$ is symmetric on $L^2(\lambda_M;\mathbb{R})$ in the sense that

$$\int_M g\mathbf{P}_t^M f\, d\lambda_M = \int_M f\mathbf{P}_t^M g\, d\lambda_M \quad \text{for all } f,g \in C\big(M;[0,\infty)\big).$$

Of course, given the preceding, we know that the non-explosion criterion in Theorem 3.64 is available to us, and so we should attempt to mimic the arguments leading to Yau's non-explosion criterion in Theorem 5.40. However, we will defer such matters until we have introduced the machinery developed in the next chapter.

CHAPTER 8

The Bundle of Orthonormal Frames

When, in Chapter 7, we abandoned our extrinsic approach, we lost the advantage provided by the Euclidean structure of the ambient space. Perhaps most important, we no longer had the natural set of vector fields (cf. (5.2)) $\{D_{\xi}^M : \xi \in \mathbb{R}^N\}$ on M: vector fields in terms of which most of the important Riemannian objects have simple expressions when M is given the Riemannian structure which it inherits from $\mathbb{R}^N$. In particular, each choice of orthonormal basis $\mathfrak{e}$ in $\mathbb{R}^N$ gave us a convenient, globally defined, collection $\left(D_1^{\mathfrak{e},M}, \dots, D_N^{\mathfrak{e},M}\right)$ of vector fields of which we made consistent and systematic use. In fact (4.33) was the starting point for everything we did in Chapters 4 and 5. On the other hand, one cannot deny that there was something slightly profligate about our use of N vector fields when one knows that $\dim(M)$ ought to suffice. This extravagance is particularly apparent in § 4.2, where we constructed Brownian motion on M by "projecting" the Brownian motion on $\mathbb{R}^N$. Indeed, as the computation in § 4.2.3 makes explicit, we were literally throwing away $N - \dim(M)$ dimensions of the randomness from the original Brownian motion in $\mathbb{R}^N$. Because our construction involved so much waste, one cannot help suspecting that it may not be the optimal. Thus, it is reasonable to suppose that a "tighter" construction exists and leads to a more compelling relationship between the original Euclidean Brownian motion and the resulting Brownian motion on the manifold. In the present chapter we will develop one such construction.

Before turning to the details, it is well to recognize what it was about the vector fields $\left(D_1^{\mathfrak{e},M}, \dots, D_N^{\mathfrak{e},M}\right)$ which made them so useful to us. After looking at the arguments used in Chapters 4 and 5, one begins to realize that a critical role was played by the following frequently exploited observation: on the one hand, the $D_k^{\mathfrak{e},M}$'s entered formulae like (4.30), (4.31), and (4.33) in a way which was invariant under change of the basis $\mathfrak{e}$, on the other hand, for a given $x \in M$, these formulae simplified when $\mathfrak{e} = (\mathbf{e}_1, \dots, \mathbf{e}_N)$ was taken so that $\{\mathbf{e}_1, \dots, \mathbf{e}_{\dim(M)}\} \subseteq T_xM$. In particular, with that choice of $\mathfrak{e}$, we had not only that $\left((D_1^{\mathfrak{e},M})_x, \dots, (D_{\dim(M)}^{\mathfrak{e},M})_x\right)$ was an orthonormal basis in T_xM while $(D_k^{\mathfrak{e},M})_x = 0$, $\dim(M) < k \le N$, but also that[1] $\nabla^M_{(D_k^{\mathfrak{e},M})_x} D_\ell^{\mathfrak{e},M} = 0$ for all $1 \le k, \ell \le N$. In the terminology of classical differential geometry, this means

[1] Recall that $\nabla_{(D_\xi^M)_x} D_\eta^M = \Pi_x^M \circ \partial\Pi_x^M(D_\xi^M)\eta = \partial\Pi_x^M(D_\xi^M) \circ (\Pi_x^M)^\perp \eta = 0$ if $\eta \in T_xM$.

that, up to first order at x, vector fields $\{D_k^{\mathfrak{c},M} : 1 \le k \le \dim(M)\}$ coincided with differentiation in the coordinate directions of a *normal* coordinate system at x (cf. §9.1. especially (9.5)) while the vector fields $\{D_k^{\mathfrak{c},M} : \dim(M) < k \le N\}$ were negligible. Thus, at least for those familiar with classical differential geometry, it is not surprising that the D_k^M's played such an important role.

Although they will, at least at first, appear to be a terrible price to pay, the contents of this chapter provide the machinery with which to write down an *intrinsic* substitute for the *extrinsic* vector fields $D_{\boldsymbol{\xi}}^M$. Thus, the reader is advised to keep this concrete goal in mind as he plows through the otherwise abstract constructions which follow.

8.1 The Bundle $\mathcal{O}(M)$

Given a d-dimensional Riemannian manifold M, the *bundle* $\mathcal{O}(M)$ *of orthonormal frames*[2] is the set of frames $\mathfrak{f} = \big(x, \mathbf{e}(x)\big)$, where $x \in M$ and $\mathbf{e}(x)$ is an orthonormal basis in T_xM. The *fiber map* $\pi : \mathcal{O}(M) \longrightarrow M$ is defined so that $\pi(\mathfrak{f}) = x$ when $\mathfrak{f} = \big(x, \mathbf{e}(x)\big)$, in which case $\mathfrak{f}$ is said to *lie above* x, $\pi^{-1}(x)$ is called the *fiber over* x, and x is the *base point below* $\mathfrak{f}$. Clearly, if $\mathfrak{f} = \big(x, \mathbf{e}(x)\big)$, then $\mathfrak{f}$ determines and is determined by the linear isometry from $\mathbb{R}^d$ onto T_xM which takes $\boldsymbol{\xi} \in \mathbb{R}^d$ into the element of T_xM whose coordinates with respect to the basis $\mathbf{e}(x)$ are the coordinates of $\boldsymbol{\xi}$ with respect to the standard orthonormal basis in $\mathbb{R}^d$. Thus, it will be convenient to identify the frame $\mathfrak{f}$ with this isometry by writing

$$\mathfrak{f}\boldsymbol{\xi} \equiv \sum_{k=1}^{d} \xi^k (E_k)_x \tag{8.1}$$
$$\text{when } \boldsymbol{\xi} = (\xi^1, \dots, \xi^d) \in \mathbb{R}^d \text{ and } \mathfrak{f} = \Big(x, \big((E_1)_x, \dots, (E_d)_x\big)\Big) \in \mathcal{O}(M).$$

In addition to the fiber map from $\mathcal{O}(M)$ onto M, the orthogonal group $O(\mathbb{R}^d)$ induces an entire group of natural maps on $\mathcal{O}(M)$ into itself. Namely, for each $O \in O(\mathbb{R}^d)$, the group of orthogonal transformations on $\mathbb{R}^d$, we have the map $R_O : \mathcal{O}(M) \longrightarrow \mathcal{O}(M)$ which is defined so that (remember that we think of $\mathfrak{f}$ as an isometry from $\mathbb{R}^d$ onto $T_{\pi\mathfrak{f}}M$)

$$(R_O\mathfrak{f})\boldsymbol{\xi} = \mathfrak{f}O\boldsymbol{\xi} \quad \text{for } \mathfrak{f} \in \mathcal{O}(M) \text{ and } \boldsymbol{\xi} \in \mathbb{R}^d. \tag{8.2}$$

Obviously, for a given $x \in \mathbb{R}^d$ and frame $\mathfrak{f}$ above x, $O \in O(\mathbb{R}^d) \longmapsto R_O\mathfrak{f} \in \mathcal{O}(M)$ provides a isomorphism from $O(\mathbb{R}^d)$ onto the fiber $\pi^{-1}(x)$ over x. In fact, if U is an open subset of M on which there exists a smoothly varying map $x \in U \longmapsto \mathbf{e}(x) \in \big(T_xM\big)^d$ such that $\mathbf{e}(x)$ is an orthonormal basis for each $x \in U$, then we get an isomorphism between $U \times O(\mathbb{R}^d)$ and $\pi^{-1}(U)$ determined by

$$(x, O) \in U \times O(\mathbb{R}^d) \longmapsto R_O\mathfrak{s}(x) \in \pi^{-1}(U) \quad \text{when } \mathfrak{s}(x) = \big(x, \mathbf{e}(x)\big).$$

[2] The reader who has no previous experience with this concept might want to consult the book [2].

Hence, since every point of M admits a open neighborhood U on which such an $x \rightsquigarrow \mathbf{e}(x)$ exists, we can turn $\mathcal{O}(M)$ into a differentiable manifold by declaring such isomorphisms to be diffeomorphisms. Equivalently, we are saying that $F : \mathcal{O}(M) \longrightarrow \mathbb{R}$ is smooth[3] if $(x, O) \in U \times O(\mathbb{R}^d) \longmapsto F\big(R_O \mathfrak{s}(x)\big) \in \mathbb{R}$ is smooth whenever $x \in U \longmapsto \mathfrak{s}(x) = \big(x, \mathbf{e}(x)\big) \in \mathcal{O}(M)$ with $x \in U \longmapsto \mathbf{e}(x) \in (T_xM)^d$ is a smooth selection of orthonormal bases on U. In this connection. we say that a map $x \rightsquigarrow \mathfrak{s}(x)$ from an open set $U \subseteq M$ to $\mathcal{O}(M)$ is a *section* of $\mathcal{O}(M)$ if it is smooth and $x = \pi \circ \mathfrak{s}(x)$ for each $x \in U$.

Now that $\mathcal{O}(M)$ has a differentiable structure, we can talk about the tangent space $T_\mathfrak{f}\mathcal{O}(M)$ to the frame $\mathfrak{f}$. Because the differentiable structure is, at least locally, simply the product of the differentiable structures on M and $O(\mathbb{R}^d)$, we know that $T_\mathfrak{f}\mathcal{O}(M)$ looks like $T_{\pi(\mathfrak{f})}M \times \mathfrak{o}(\mathbb{R}^d)$, where $\mathfrak{o}(\mathbb{R}^d)$ is the Lie algebra of skew symmetric $d \times d$-matrices. In fact, it is easy to understand the part of $T_\mathfrak{f}\mathcal{O}(M)$ which corresponds to $\mathfrak{o}(\mathbb{R}^d)$. Namely, $\mathfrak{o}(\mathbb{R}^d)$ should parameterize the *vertical* subspace $V_\mathfrak{f}\mathcal{O}(M)$ of $\mathfrak{X}_\mathfrak{f} \in T_\mathfrak{f}\mathcal{O}(M)$ which are tangent to motion which is purely up and down the fiber $\pi^{-1}\big(\pi(\mathfrak{f})\big)$ and does not move the base point. Equivalently, $\mathfrak{X}_\mathfrak{f}$ is *vertical* if $\pi_*\mathfrak{X}_\mathfrak{f} = 0$. To parameterize $V_\mathfrak{f}\mathcal{O}(M)$ by $\mathfrak{o}(\mathbb{R}^d)$, let $\mathfrak{X}_\mathfrak{f} \in V_\mathfrak{f}\mathcal{O}(M)$ be given, and choose $t \in \mathbb{R} \longmapsto \mathfrak{p}(t) \in \mathcal{O}(M)$ to be a smooth curve such that $\mathfrak{p}(0) = \mathfrak{f}$, $\dot{\mathfrak{p}}(0) = \mathfrak{X}_\mathfrak{f}$, and $\pi \circ \mathfrak{p}(t) = \pi\mathfrak{f}$ for all t. In particular, the last property means that $\mathfrak{p}(t) = R_{O(t)}\mathfrak{f}$ when $t \in \mathbb{R} \longmapsto O(t) \equiv \mathfrak{f}^{-1} \circ \mathfrak{p}(t) \in O(\mathbb{R}^d)$. Hence, $a \equiv \dot{O}(0)$ is an element of $\mathfrak{o}(\mathbb{R}^d)$ and $\mathfrak{X}_\mathfrak{f} = \lambda(a)_\mathfrak{f}$, where $\lambda(a)$ is the vector field on $\mathcal{O}(M)$ such that

$$\lambda(a)_\mathfrak{f} = \frac{d}{ds} R_{e^{sa}}\mathfrak{f}\Big|_{s=0}. \tag{8.3}$$

In other words,

$$V_\mathfrak{f}\mathcal{O}(M) \equiv \big\{\mathfrak{X}_\mathfrak{f} : \pi_*\mathfrak{X}_\mathfrak{f} = 0\big\} = \big\{\lambda(a)_\mathfrak{f} : a \in \mathfrak{o}(\mathbb{R}^d)\big\}. \tag{8.4}$$

Clearly, the distribution $\mathfrak{f} \rightsquigarrow V_\mathfrak{f}\mathcal{O}(M)$ is integrable: for each $\mathfrak{f} \in \mathcal{O}(M)$, the fiber on which $\mathfrak{f}$ lies is the maximal integral manifold of this distribution passing through $\mathfrak{f}$. In fact, of course, these fibers are just copies of $O(\mathbb{R}^d)$. Moreover, because $R_O \circ R_{O'} = R_{O'O}$, it is an easy matter to check that the vector field $\lambda(a)$ acts on each fiber as the representative of the right-invariant vector field on $O(\mathbb{R}^d)$ determined by $a \in o(\mathbb{R}^d)$. In particular, one finds that

$$(R_O)_*\lambda(a)_\mathfrak{f} = \lambda\big(O^\top a O\big)_{R_O\mathfrak{f}} \quad \text{and} \quad \big[\lambda(a), \lambda(a')\big]_\mathfrak{f} = \lambda\big([a, a']\big)_\mathfrak{f}. \tag{8.5}$$

8.1.1. The Riemannian Connection and the Horizontal Subspace. As we have just seen, as soon as M has a Riemannian structure, both $\mathcal{O}(M)$ and the vertical subspaces are completely canonical. On the other hand, we are still left with the problem of designating a subspace $H_\mathfrak{f}\mathcal{O}(M)$ of $T_\mathfrak{f}\mathcal{O}(M)$ which will

[3] Here again we are assuming that the reader is familiar with the fact the $O(\mathbb{R}^d)$ is a differentiable manifold.

correspond to motion which is *horizontal* in the sense that it involves *no change* of basis. Of course, until we know how to compare bases at two different points of M, there is no hope of our recognizing whether or not change has occurred. Hence, it is at this point that the Riemannian connection gets encoded into $\mathcal{O}(M)$. To be precise, given a smooth curve $t \in [a,b] \longmapsto p(t) \in M$, a *horizontal lift* of p to $\mathcal{O}(M)$ is a curve $t \in [a,b] \longmapsto \mathfrak{p}(t) = \big(p(t), \mathbf{e}(t)\big) \in \mathcal{O}(M)$ where, for each pair $a \le t_1 < t_2 \le b$, (cf. §7.2.3)

$$\mathbf{e}(t_2) = \mathcal{T}_{p\restriction[t_1,t_2]}\mathbf{e}(t_1).$$

That is, the kth element of the basis $\mathbf{e}(t_2)$ is obtained from the kth element of $\mathbf{e}(t_1)$ by parallel transport along $t \in [t_1,t_2] \longmapsto p(t) \in M$. Equivalently, $t \rightsquigarrow \mathfrak{p}(t)$ is a horizontal lift of $t \rightsquigarrow p(t)$ if $p(t) = \pi \circ \mathfrak{p}(t)$ and $\frac{D\mathfrak{p}(t)\boldsymbol{\xi}}{dt} = 0$ for every $\boldsymbol{\xi} \in \mathbb{R}^d$. In particular, for each $\mathfrak{f} \in \pi^{-1}\big(p(a)\big)$, there is precisely one horizontal lift $\mathfrak{p}$ of p which passes through $\mathfrak{f}$ at $t = a$. Conversely, we will say that a smooth curve $t \rightsquigarrow \mathfrak{p}(t)$ in $\mathcal{O}(M)$ is *horizontal* if it is the horizontal lift of $t \rightsquigarrow \pi \circ \mathfrak{p}(t)$. Finally, given $x \in M$, $X_x \in T_xM$, and $\mathfrak{f} \in \pi^{-1}(x)$, we say that $\mathfrak{X}_\mathfrak{f} \in T_\mathfrak{f}\mathcal{O}(M)$ is a *horizontal lift* of X_x to $\mathfrak{f}$ if $\mathfrak{X}_\mathfrak{f} = \dot{\mathfrak{p}}(0)$ where $t \rightsquigarrow \mathfrak{p}(t)$ is a horizontal curve with the properties that $\mathfrak{p}(0) = \mathfrak{f}$ and $\pi_*\dot{\mathfrak{p}}(0) = X_x$.

8.6 LEMMA. *If $\mathfrak{p} : [a,b] \longrightarrow \mathcal{O}(M)$ is horizontal, then, for each $O \in O(\mathbb{R}^d)$, $t \in [a,b] \longmapsto R_O \circ \mathfrak{p}(t) \in \mathcal{O}(M)$ is also horizontal. Moreover, for each $x \in M$, $X_x \in T_xM$, and $\mathfrak{f} \in \pi^{-1}(x)$, there is precisely one horizontal lift $\mathfrak{H}(X_x)_\mathfrak{f} \in T_\mathfrak{f}\mathcal{O}(M)$ of X_x to $\mathfrak{f}$. Furthermore, the map $\mathfrak{H}$ is linear and $O(\mathbb{R}^d)$-equivariant in the sense that, for each $x \in M$ and $\mathfrak{f} \in \pi^{-1}(x)$, $X_x \in T_xM \longmapsto \mathfrak{H}(X_x)_\mathfrak{f} \in T_\mathfrak{f}\mathcal{O}(M)$ is linear and*

$$\mathfrak{H}(X_x)_{R_O\mathfrak{f}} = (R_O)_*\mathfrak{H}(X_x)_\mathfrak{f} \quad \text{for all } O \in O(\mathbb{R}^d). \tag{8.7}$$

Finally, if X is a smooth vector field on an open subset U of M, then $\mathfrak{f} \in \pi^{-1}(U) \longmapsto \mathfrak{H}(X_{\pi\mathfrak{f}})_\mathfrak{f} \in T_\mathfrak{f}\mathcal{O}(M)$ is also a smooth vector field.

PROOF: The first assertion is evident since if $\mathfrak{p}$ is horizontal then $\frac{DR_O\mathfrak{p}(t)\boldsymbol{\xi}}{dt} = \frac{D\mathfrak{p}(t)O\boldsymbol{\xi}}{dt} = 0$ for any $O \in O(\mathbb{R}^d)$ and $\boldsymbol{\xi} \in \mathbb{R}^d$.

In proving the rest of the lemma, let U be an open set in M on which there exists a smooth section

$$x \rightsquigarrow \mathfrak{s}(x) = \big(x, \mathbf{e}(x)\big) = \Big(x, \big((E_1)_x, \dots, (E_d)_x\big)\Big).$$

Given a smooth curve $p : [a,b] \longrightarrow U$, let $t \in [a,b] \longmapsto O_p(t) \in O(\mathbb{R}^d)$ be the associated curve determined by (7.17), and observe that $\mathfrak{p}$ is a horizontal lift of p if and only if

$$\mathfrak{p}(t) = R_{\mathfrak{s}(p(a))^{-1}\circ\mathfrak{p}(a)} \circ R_{O_p(t)}\mathfrak{s}\big(p(t)\big), \quad t \in [a,b]. \tag{8.8}$$

Thus, if, for $x \in U$ and $X_x \in T_xM$, (cf. (7.12))

$$A(X_x) = \big(\big(A(X_x)^k_j\big)\big) \quad \text{where } A(X_x)^k_j \equiv \sum_{i=1}^d \langle X_x, (E_i)_x\rangle \big[\Gamma^k_{i,j}(\mathbf{e})\big](x), \tag{8.9}$$

and if $\mathfrak{S}(X_x)_{\mathfrak{s}(x)}$ is the element of $T_{\mathfrak{s}(x)}\mathcal{O}(M)$ determined by $\mathfrak{S}(X_x)_{\mathfrak{s}(x)}F = \mathfrak{s}_* X_x F = X_x F \circ \mathfrak{s}$, then for any horizontal lift $\mathfrak{p}$ of p, (cf. (8.8))

$$\dot{\mathfrak{p}}(t) = \left(R_{\mathfrak{s}(p(a))^{-1}\circ\mathfrak{p}(a)}\right)_*\Big(\mathfrak{S}\big(\dot{p}(t)\big)_{\mathfrak{s}(p(t))} - \lambda\big(A(\dot{p}(t))\big)_{\mathfrak{s}(p(t))}\Big).$$

In particular, this proves that if $x \in U$ and $\mathfrak{f} \in \pi^{-1}(x)$, then the one and only horizontal lift of X_x to $\mathfrak{f}$ is given by

$$\mathfrak{H}(X_x)_{\mathfrak{f}} = \left(R_{\mathfrak{s}(x)^{-1}\circ\mathfrak{f}}\right)_*\Big(\mathfrak{S}(X_x)_{\mathfrak{s}(x)} - \lambda\big(A(X_x)\big)_{\mathfrak{s}(x)}\Big). \tag{8.10}$$

Clearly, all the other asserted properties of $\mathfrak{H}$ follow from (8.10) □

Now, given $\mathfrak{f} \in \mathcal{O}(M)$, we take the *horizontal subspace* at $\mathfrak{f}$ to be the space

$$H_{\mathfrak{f}}\mathcal{O}(M) \equiv \left\{\mathfrak{H}\big(X_{\pi\mathfrak{f}}\big)_{\mathfrak{f}} : X_{\pi\mathfrak{f}} \in T_{\pi\mathfrak{f}}M\right\} \tag{8.11}$$

and say that the vector $\mathfrak{X}_{\mathfrak{f}} \in T_{\mathfrak{f}}\mathcal{O}(M)$ is *horizontal* if $\mathfrak{X}_{\mathfrak{f}} \in H_{\mathfrak{f}}\mathcal{O}(M)$. Because $\mathfrak{H}(X_{\pi\mathfrak{f}})_{\mathfrak{f}} \in V_{\mathfrak{f}}\mathcal{O}(M)$ only if $X_{\pi\mathfrak{f}} = \mathbf{0}$, we know that, at each $\mathfrak{f} \in \mathcal{O}(M)$, $T_{\mathfrak{f}}\mathcal{O}(M) = H_{\mathfrak{f}}\mathcal{O}(M) \oplus V_{\mathfrak{f}}\mathcal{O}(M)$.

To highlight the virtues of this decomposition, we introduce[4] the *canonical vector fields* $\mathfrak{E}(\boldsymbol{\xi})$, $\boldsymbol{\xi} \in \mathbb{R}^d$, the *solder form* ϕ, and the *connection 1-form* ω. Namely, for each $\boldsymbol{\xi} \in \mathbb{R}^d$, $\mathfrak{E}(\boldsymbol{\xi})$ is the vector field on $\mathcal{O}(M)$ such that

$$\mathfrak{E}(\boldsymbol{\xi})_{\mathfrak{f}} \in H_{\mathfrak{f}}\mathcal{O}(M) \text{ and } \pi_*\mathfrak{E}(\boldsymbol{\xi})_{\mathfrak{f}} = \mathfrak{f}\boldsymbol{\xi} \quad \text{for each } \mathfrak{f} \in \mathcal{O}(M). \tag{8.12}$$

Equivalently, $\mathfrak{E}(\boldsymbol{\xi})_{\mathfrak{f}} = \mathfrak{H}(\mathfrak{f}\boldsymbol{\xi})_{\mathfrak{f}}$, and therefore (8.10) says that

$$\mathfrak{E}(\boldsymbol{\xi})_{\mathfrak{f}} = \left(R_{\mathfrak{s}(\pi\mathfrak{f})^{-1}\circ\mathfrak{f}}\right)_*\Big(\mathfrak{S}(\mathfrak{f}\boldsymbol{\xi})_{\mathfrak{s}(\pi\mathfrak{f})} - \lambda\big(A(\mathfrak{f}\boldsymbol{\xi})\big)_{\mathfrak{s}(\pi\mathfrak{f})}\Big) \quad \text{for } \mathfrak{f} \in \pi^{-1}(U), \tag{8.13}$$

from which it is clear that $\mathfrak{E}(\boldsymbol{\xi})$ is indeed a smooth vector field on $\mathcal{O}(M)$. In addition, from (8.7), we have that

$$(R_O)_*\mathfrak{E}(\boldsymbol{\xi})_{\mathfrak{f}} = \mathfrak{E}\big(O^\top\boldsymbol{\xi}\big)_{R_O\mathfrak{f}}. \tag{8.14}$$

Also, it will be useful to know that

$$\big[\lambda(a), \mathfrak{E}(\boldsymbol{\xi})\big] = \mathfrak{E}(a\boldsymbol{\xi}), \quad a \in o(\mathbb{R}^d) \text{ and } \boldsymbol{\xi} \in \mathbb{R}^d. \tag{8.15}$$

To prove this, use (8.14) to see that

$$\lambda(a)_{\mathfrak{f}}\mathfrak{E}(\boldsymbol{\xi})F = \frac{d}{ds}\mathfrak{E}\left(e^{sa}\boldsymbol{\xi}\right)F \circ R_{e^{sa}}\bigg|_{s=0} = \mathfrak{E}(a\boldsymbol{\xi})_{\mathfrak{f}}F + \mathfrak{E}(\boldsymbol{\xi})_{\mathfrak{f}}\lambda(a)F.$$

[4] The reader who followed my earlier advise and "consulted" [2] should be warnad that I have *reversed* the rôles that the symbols ϕ and ω play in that book.

Next, define the $\mathbb{R}^d$-valued 1-form ϕ so that

$$(8.16) \qquad \phi(\mathfrak{X}_\mathfrak{f}) = \mathfrak{f}^{-1}(\pi_* \mathfrak{X}_\mathfrak{f}\mathfrak{f}) \quad \text{for } \mathfrak{f} \in \mathcal{O}(M) \text{ and } X_\mathfrak{f} \in T_\mathfrak{f}\mathcal{O}(M).$$

Alternatively, $\phi(\mathfrak{X}_\mathfrak{f}) \in \mathbb{R}^d$ is determined by $\mathfrak{E}(\phi(\mathfrak{X}_\mathfrak{f}))_\mathfrak{f} = \mathfrak{H}(\pi_*\mathfrak{X}_\mathfrak{f})_\mathfrak{f}$. At the same time, recall that $\mathfrak{X}_\mathfrak{f} - \mathfrak{E}(\phi(\mathfrak{X}_\mathfrak{f}))_\mathfrak{f} \in V_\mathfrak{f}\mathcal{O}(M)$, and determine the $o(\mathbb{R}^d)$-valued 1-form ω on $\mathcal{O}(M)$ by (cf. (8.3))

$$(8.17) \qquad \lambda(\omega(\mathfrak{X}_\mathfrak{f}))_\mathfrak{f} = \mathfrak{X}_\mathfrak{f} - \mathfrak{E}(\phi(\mathfrak{X}_\mathfrak{f}))_\mathfrak{f} \quad \text{for } \mathfrak{f} \in \mathcal{O}(M) \text{ and } \mathfrak{X}_\mathfrak{f} \in T_\mathfrak{f}\mathcal{O}(M).$$

Obviously,

$$(8.18) \qquad \mathfrak{X}_\mathfrak{f} = \mathfrak{E}(\phi(\mathfrak{X}_\mathfrak{f}))_\mathfrak{f} + \lambda(\omega(\mathfrak{X}_\mathfrak{f}))_\mathfrak{f} \in H_\mathfrak{f}\mathcal{O}(M) \oplus V_\mathfrak{f}\mathcal{O}(M)$$

gives the decomposition of a vector $\mathfrak{X}_\mathfrak{f} \in T_\mathfrak{f}\mathcal{O}(M)$ into its *horizontal* and its *vertical* parts. In particular, by (8.14) and the first part of (8.5),

$$(R_O)_*\mathfrak{X}_\mathfrak{f} = \mathfrak{E}(O^\top \phi(\mathfrak{X}_\mathfrak{f}))_{R_O\mathfrak{f}} + \lambda(O^\top \omega(\mathfrak{X}_\mathfrak{f})O)_{R_O\mathfrak{f}},$$

and so

$$(8.19) \qquad \phi((R_O)_*\mathfrak{X}_\mathfrak{f}) = O^\top \phi(\mathfrak{X}_\mathfrak{f}) \quad \text{and} \quad \omega((R_O)_*\mathfrak{X}_\mathfrak{f}) = O^\top \omega(\mathfrak{X}_\mathfrak{f})O.$$

Although it is obvious that ϕ is smooth, the smoothness of ω requires some checking. For this purpose, let $U \subseteq M$ be an non-empty open subset on which there is a smooth section $x \in U \longmapsto \mathfrak{s}(x) \in \pi^{-1}x$. Given $x \in U$ and $\mathfrak{X}_{\mathfrak{s}(x)} \in T_{\mathfrak{s}(x)}\mathcal{O}(M)$, choose a smooth $\mathfrak{q} : [0,1] \longrightarrow \pi^{-1}(U)$ so that $\mathfrak{q}(0) = \mathfrak{s}(x)$ and $\dot{\mathfrak{q}}(0) = \mathfrak{X}_{\mathfrak{s}(x)}$. Next, set $p(t) = \pi \circ \mathfrak{q}(t)$ and (cf. (8.8)) $\mathfrak{p}(t) = R_{O_p(t)}\mathfrak{s}(p(t))$. Then $\pi \circ \mathfrak{p}(t) = p(t) = \pi \circ \mathfrak{q}(t)$, $\mathfrak{p}$ is horizontal, and $\mathfrak{q}(t) = R_{\mathfrak{p}(t)^{-1}\circ\mathfrak{q}(t)}\mathfrak{p}(t)$. Thus,

$$\omega(\mathfrak{X}_{\mathfrak{s}(x)}) = \frac{d}{dt}\mathfrak{p}(t)^{-1} \circ \mathfrak{q}(t)\Big|_{t=0}.$$

But if $G : \pi^{-1}(U) \longrightarrow O(\mathbb{R}^d)$ is given by

$$G(\mathfrak{f}) = \mathfrak{s}(\pi(\mathfrak{f}))^{-1} \circ \mathfrak{f}, \quad \mathfrak{f} \in \pi^{-1}(U),$$

then $\mathfrak{p}(t)^{-1} \circ \mathfrak{q}(t) = O_p(t)^\top \circ G(\mathfrak{q}(t))$, and therefore (cf. (8.9))

$$\omega(\mathfrak{X}_{\mathfrak{s}(x)}) = A(\pi_*\mathfrak{X}_{\mathfrak{s}(x)}) + \mathfrak{X}_{\mathfrak{s}(x)}G.$$

Finally, after combining this with (8.19), we see that

$$\begin{aligned}\omega(\mathfrak{X}_\mathfrak{f}) = \omega((R_{G(\mathfrak{f})})_* \circ (R_{G(\mathfrak{f})^\top})_*\mathfrak{X}_{\mathfrak{s}(\pi\mathfrak{f})}) &= G(\mathfrak{f})^\top \omega((R_{G(\mathfrak{f})^\top})_*\mathfrak{X}_{\mathfrak{s}(\pi\mathfrak{f})})G(\mathfrak{f}) \\ &= G(\mathfrak{f})^\top \Big(A(\pi_*\mathfrak{X}_\mathfrak{f}) + \mathfrak{X}_\mathfrak{f}(G \circ R_{G(\mathfrak{f})^\top})\Big)G(\mathfrak{f})\end{aligned}$$

for all $\mathfrak{f} \in \pi^{-1}(U)$ and $\mathfrak{X}_\mathfrak{f} \in T_\mathfrak{f}\mathcal{O}(M)$; and clearly this proves the required smoothness of ω.

As a dividend of these considerations, we now have the following prescription for horizontally lifting curves and indication of what to do once we have lifted them.

8.20 LEMMA. *Let* $p : [a,b] \longrightarrow M$ be a smooth curve and $\mathfrak{f} \in \pi^{-1}(p(a))$. *Then the curve* $\mathfrak{p} : [a,b] \longrightarrow \mathcal{O}(M)$ *determined by*

$$\dot{\mathfrak{p}}(t) = \mathfrak{E}\big(\mathfrak{p}(t)^{-1}\dot{p}(t)\big)_{\mathfrak{p}(t)} \quad \text{with } \mathfrak{p}(a) = \mathfrak{f}$$

is the one and only horizontal lift of p *which passes through* $\mathfrak{f}$ *at* $t = a$. *Moreover, if* $t \in [a,b] \longmapsto Y(t) \in T_{p(t)}M$ is a smooth vector field along p, *then*

$$\mathfrak{f}\mathfrak{p}(t)^{-1}Y(t) = \mathcal{T}^{-1}_{p\restriction[a,t]}Y(t), \quad t \in [a,b],$$

and so

$$\frac{DY}{dt}(a) = \mathfrak{f}\left(\frac{d}{dt}\mathfrak{p}(t)^{-1}Y(t)\right)\bigg|_{t=a}.$$

In particular, if Y is a vector field on M and $\Xi_Y : \mathcal{O}(M) \longrightarrow \mathbb{R}^d$ is *defined* so *that*

$$\Xi_Y(\mathfrak{f}) = \mathfrak{f}^{-1}Y_{\pi(\mathfrak{f})} \quad \text{for } \mathfrak{f} \in \mathcal{O}(M), \tag{8.21}$$

then,

$$\Xi_{\nabla^M_{\mathfrak{f}\xi}Y}(\mathfrak{f}) = \mathfrak{f}^{-1}\nabla^M_{\mathfrak{f}\xi}Y = \mathfrak{E}(\xi)_{\mathfrak{f}}\Xi_Y \quad \text{for all } \xi \in \mathbb{R}^d \text{ and } \mathfrak{f} \in \mathcal{O}(M). \tag{8.22}$$

Hence, if $\mathfrak{f} \in \pi^{-1}x$, $\mathfrak{X}_{\mathfrak{f}} \in T_{\mathfrak{f}}\mathcal{O}(M)$, and $X_x = \pi_*\mathfrak{X}_{\mathfrak{f}}$, *then*

$$\mathfrak{f}^{-1}\nabla_{X_x}Y = \mathfrak{X}_{\mathfrak{f}}\Xi_Y + \omega(\mathfrak{X}_{\mathfrak{f}})\Xi_Y(\mathfrak{f}). \tag{8.23}$$

PROOF: To handle the first assertion, observe that, because $\mathfrak{E}\big(\mathfrak{f}^{-1}\dot{p}(t)\big)_{\mathfrak{f}} = \mathfrak{H}\big(\dot{p}(t)\big)_{\mathfrak{f}}$, the given differential equation determines a unique curve for each $\mathfrak{f} \in \pi^{-1}\big(p(a)\big)$. Thus, it suffices to note that if $\mathfrak{p}$ is a horizontal lift of p, then, because $\dot{\mathfrak{p}}(t)$ will be horizontal, $\dot{\mathfrak{p}}(t) = \mathfrak{E}\big(\mathfrak{p}(t)^{-1}\dot{p}(t)\big)$ for all $t \in [a,b]$.

Turning to the second part, simply note that

$$\mathcal{T}_{p\restriction[a,t]}\mathfrak{f}\big(\mathfrak{p}(t)^{-1}Y(t)\big) = \mathfrak{p}(t)\big(\mathfrak{p}(t)^{-1}Y(t)\big) = Y(t).$$

and so everything except (8.22) and (8.23) follows immediately from Lemma 7.16 and (5.16). To prove (8.22), choose $\delta > 0$ so that there exists a solution to $\dot{\mathfrak{p}}(t) = \mathfrak{E}(\xi)_{\mathfrak{p}(t)}$ with $\mathfrak{p}(0) = \mathfrak{f}$ for $|t| \le \delta$. Then, $\mathfrak{p}$ is the horizontal lift of $p \equiv \pi \circ \mathfrak{p}$ with $\mathfrak{p}(0) = \mathfrak{f}$, and so, by the preceding,

$$\mathfrak{f}\big(\mathfrak{E}(\xi)_{\mathfrak{f}}\Xi_Y\big) = \frac{d}{dt}\mathfrak{f}\big(\mathfrak{p}(t)^{-1}Y_{p(t)}\big)\Big|_{t=0} = \frac{DY_{p(t)}}{dt}\Big|_{t=0} = \nabla^M_{\mathfrak{f}\xi}Y.$$

Finally, given (8.22), (8.23) comes from setting $\xi = \mathfrak{f}^{-1}X_x$, $a = \omega(\mathfrak{X}_{\mathfrak{f}})$, and writing $\mathfrak{E}(\xi)_{\mathfrak{f}} = \mathfrak{X}_{\mathfrak{f}} - \lambda(a)_{\mathfrak{f}}$. Indeed, (8.22) then says that $\mathfrak{f}^{-1}\nabla_{X_x}Y = \mathfrak{X}_{\mathfrak{f}}\Xi_Y - \lambda(a)_{\mathfrak{f}}\Xi_Y$. Hence, since $R_O\Xi_Y = O^{\top}\Xi_Y$ and therefore

$$\lambda(a)_{\mathfrak{f}}\Xi_Y = \lim_{s\searrow 0} s^{-1}\Big(\Xi_Y\big(R_{e^{sa}}f\big) - \Xi_Y(\mathfrak{f})\Big) = \lim_{s\searrow 0} s^{-1}\big(e^{-sa} - I\big)\Xi_Y(\mathfrak{f}) = -a\Xi_Y(\mathfrak{f}).$$

(8.23) follows.

8.1.2. Rolling, Geodesics, and Completeness. A fundamental construction in differential geometry takes a smooth curve $t \in [a,b] \longmapsto X(t) \in T_xM$ which starts at the origin in the tangent space to a point $x \in M$ and "rolls" it onto M. Namely, one *rolls* $t \rightsquigarrow X(t)$ onto a curve p in M by the prescription:

$$p(a) = x \text{ and } \dot{p}(t) = \mathcal{T}_{p\restriction[a,t]}\dot{X}(t) \quad \text{for} \quad t \in [a,b]. \tag{8.24}$$

The idea is that the rolled curve $t \rightsquigarrow p(t)$ should be the curve in M which starts at x and proceeds "with the same velocity" as $t \rightsquigarrow X(t)$, where "same" means the parallel transport of $\dot{X}(t)$[5] along that part of the curve which has already been created during $[a,t]$. To see how one might go about actually constructing the roll of $t \rightsquigarrow X(t)$ onto M, suppose that p is a solution to (8.24), and let $\mathfrak{p}$ be horizontal lift of p with $\mathfrak{p}(a) = \mathfrak{f} \in \pi^{-1}(p(a))$. Then, by Lemma 8.20,

$$\dot{\mathfrak{p}}(t) = \mathfrak{E}\big(\mathfrak{f}^{-1}\dot{X}(t)\big)_{\mathfrak{p}(t)} \quad \text{with } \mathfrak{p}(a) = \mathfrak{f}, \tag{8.25}$$

a fact which demonstrates that $\mathfrak{p}$ is uniquely determined. Conversely, if $\mathfrak{p}$ solves (8.25) and $p = \pi \circ \mathfrak{p}$, then

$$\dot{p}(t) = \mathfrak{p}(t)\mathfrak{f}^{-1}X(t) = \mathcal{T}_{p\restriction[a,t]}X(t),$$

and so p is the one and only solution to (8.24).

A particularly important case of the rolling construction is the one when $t \rightsquigarrow X(t)$ is linear with velocity X_x. In this case, (8.24) is equivalent to

$$\frac{D\dot{p}}{dt}(t) = \mathbf{0}, \quad t \in [a,b], \quad \text{with } p(a) = x \text{ and } \dot{p}(a) = X_x. \tag{8.26}$$

In the terminology of classical differential geometry, (8.26) is saying that p is the *geodesic* during $[a,b]$ with initial position x and initial velocity X_x. Notice that $\frac{D\dot{p}}{dt}(t) = \mathbf{0}$ has the interpretation that *p has zero acceleration.*

Clearly, as long as a solution to (8.25) exists, it is unique. Thus, the only question is that of lifetime: for how long does a solution exist? We will say that M is *complete Riemannian manifold* if, for every $x \in M$ and $X_x \in T_xM$, there is a geodesic p during $[0,\infty)$ starting from x with velocity X_x. Equivalently, M is complete if, for each $\boldsymbol{\xi} \in \mathbb{R}^d$, the canonical vector field $\mathfrak{E}(\boldsymbol{\xi})$ is complete. In particular, the elementary theory of ordinary differential equations allows us to conclude that completeness of M guarantees that, for every $\mathfrak{f} \in \mathcal{O}(M)$ and every smooth $t \in [0,\infty) \longrightarrow \mathbf{w}(t) \in \mathbb{R}^d$, there is precisely one $t \in [0,\infty) \longmapsto \mathfrak{p}(t,\mathfrak{f},\mathbf{w}) \in \mathcal{O}(M)$ satisfying

$$\dot{\mathfrak{p}}(t,\mathfrak{f},\mathbf{w}) = \mathfrak{E}(\dot{\mathbf{w}}(t))_{\mathfrak{p}(t,\mathfrak{f},\mathbf{w})} \quad \text{with } \mathfrak{p}(0,\mathfrak{f},\mathbf{w}) = \mathfrak{f}. \tag{8.27}$$

[5] We are interpreting $\dot{X}(t)$ as an element of T_xM via the natural identification $T_xM \simeq T_{X(t)}(T_xM)$.

Equivalently, $\mathfrak{p}(\cdot, \mathfrak{f}, \mathbf{w})$ is *the horizontal lift starting at* $\mathfrak{f}$ *of the curve obtained by rolling* $t \rightsquigarrow \mathfrak{f}\mathbf{w}(t)$. In fact, for each smooth $\mathbf{w}$, $(t, \mathfrak{f}) \in [0, \infty) \times \mathcal{O}(M) \longmapsto \mathfrak{p}(t, \mathfrak{f}, \mathbf{w}) \in \mathcal{O}(M)$ will be a smooth map which satisfies the flow property

$$\begin{aligned} \mathfrak{p}(s+t, \mathfrak{f}, \mathbf{w}) &= \mathfrak{p}\big(t, \mathfrak{p}(s, \mathfrak{f}, \mathbf{w}), \delta_s \mathbf{w}\big), \\ &\quad \text{where } \big[\delta_s \mathbf{w}\big](\tau) = \mathbf{w}(s+\tau) - \mathbf{w}(s). \end{aligned} \tag{8.28}$$

8.1.3. Canonical Vector Fields and the Laplacian. From our point of view, the primary justification for all these considerations is the fact that, for any choice $\mathfrak{e} = (\mathbf{e}_1, \dots, \mathbf{e}_d)$ of orthonormal basis in $\mathbb{R}^d$,

$$\Delta_M = \pi_* \Delta_B \quad \text{where } \Delta_B \equiv \sum_{k=1}^d \mathfrak{E}(\mathbf{e}_k)^2. \tag{8.29}$$

To prove (8.29), first note that, by (7.20) and (5.11),

$$H_x^M f\, X_x = \frac{d}{dt} \mathcal{T}_{p\restriction[0,t]}^{-1} \mathrm{grad}_{p(t)}^M f \Big|_{t=0}$$

if p is a smooth curve with $p(0) = x$ and $\dot{p}(0) = X_x$. Hence, if $Y_x \in T_x M$, then

$$\big\langle Y_x, H_x^M f\, X_x \big\rangle = \frac{d}{dt} \mathcal{T}_{p\restriction[0,t]} Y_x\, f \Big|_{t=0}$$

for any such curve p. Now let $\boldsymbol{\xi}, \boldsymbol{\eta} \in \mathbb{R}^d$ and $\mathfrak{f} \in \mathcal{O}(M)$ be given. Then, even if M is not complete, we can find a $\delta > 0$ and a $t \in [0, \delta) \longmapsto \mathfrak{p}(t) \in \mathcal{O}(M)$ such that $\mathfrak{p}(0) = \mathfrak{f}$ and $\dot{\mathfrak{p}}(t) = \mathfrak{E}(\boldsymbol{\xi})_{\mathfrak{p}(t)}$ for $t \in [0, \delta)$. Thus, with $p = \pi \circ \mathfrak{p}$,

$$\mathfrak{E}(\boldsymbol{\xi})_{\mathfrak{f}} \circ \mathfrak{E}(\boldsymbol{\eta}) f \circ \pi = \frac{d}{dt} \mathfrak{p}(t) \boldsymbol{\eta}\, f \Big|_{t=0} = \frac{d}{dt} \mathcal{T}_{p\restriction[0,t]} \mathfrak{f}\boldsymbol{\eta}\, f \Big|_{t=0} = \big\langle \mathfrak{f}\boldsymbol{\eta}, H_{\pi\mathfrak{f}}^M f\, \mathfrak{f}\boldsymbol{\xi} \big\rangle$$

and so

$$\mathfrak{E}(\boldsymbol{\xi})_{\mathfrak{f}} \circ \mathfrak{E}(\boldsymbol{\eta}) \big(f \circ \pi\big) = \big\langle \mathfrak{f}\boldsymbol{\eta}, H_{\pi\mathfrak{f}}^M f\, \mathfrak{f}\boldsymbol{\xi} \big\rangle. \tag{8.30}$$

Clearly (8.29) follows immediately from the conjunction of (7.23) with (8.30).

We will call the operator Δ_B, which was introduced in (8.29), *Bochner's Laplacian.* It will play an important role in our treatment of Brownian motion in the present context.

8.1.4. A Measure on $\mathcal{O}(M)$. As our final topic in this cursory introduction to the bundle of orthonormal frames, we will put a "natural" measure $\lambda_{\mathcal{O}(M)}$ on $\mathcal{O}(M)$. Namely, use $\lambda_{\mathcal{O}(\mathbb{R}^d)}$ to denote the normalized Haar measure on $\mathcal{O}(\mathbb{R}^d)$. We want to define $\lambda_{\mathcal{O}(M)}$ so that if $x \in U \longmapsto \mathfrak{s}(x) \in \mathcal{O}(M)$ is a section on the open set $U \subseteq M$, then

$$\int_{\pi^{-1}(U)} F\, d\lambda_{\mathcal{O}(M)} = \int_U \left(\int_{\mathcal{O}(\mathbb{R}^d)} F\big(R_O \mathfrak{s}(x)\big)\, \lambda_{\mathcal{O}(\mathbb{R}^d)}(dO) \right) \lambda_M(dx) \tag{8.31}$$

for $F \in C\big(\mathcal{O}(M);[0,\infty)\big)$. To see that this is possible, suppose that $x \in U' \longmapsto \mathfrak{s}'(x) \in \mathcal{O}(M)$ is a second section, and use the invariance of $\lambda_{O(\mathbb{R}^d)}$ to check that

$$\int_{U\cap U'}\left(\int_{O(\mathbb{R}^d)}\Big(F\big(R_O\mathfrak{s}(x)\big) - F\big(R_O\mathfrak{s}'(x)\big)\Big)\,\lambda_{O(\mathbb{R}^d)}(dO)\right)\lambda_M(dx) = 0.$$

To justify the claim that $\lambda_{\mathcal{O}(M)}$ is natural, we will now show that, for each $\boldsymbol{\xi} \in \mathbb{R}^d$,

$$\mathfrak{E}(\boldsymbol{\xi})^{*\lambda_{\mathcal{O}(M)}} = -\mathfrak{E}(\boldsymbol{\xi}). \tag{8.32}$$

In particular, this will certainly show that Bochner's Laplacian Δ_B is symmetric on $C_c^\infty\big(\mathcal{O}(M);\mathbb{R}\big)$ in $L^2(\lambda_{\mathcal{O}(M)};\mathbb{R})$. To prove (8.32), it suffices to show that if

$$x \in U \longmapsto \mathfrak{s}(x) = \big(x, \mathbf{e}(x)\big) = \Big(x, \big((E_1)_x, \dots, (E_d)_x\big)\Big) \in \mathcal{O}(M)$$

is a smooth section and $F \in C_c^\infty\big(\pi^{-1}(U);\mathbb{R}\big)$, then the $\lambda_{\mathcal{O}(M)}$-integral of $\mathfrak{E}(\boldsymbol{\xi})F$ on $\pi^{-1}(U)$ vanishes. To this end, note that, by (8.13) and (8.14),

$$\begin{aligned}\mathfrak{E}(\boldsymbol{\xi})_{R_O\mathfrak{s}(x)}F &= \mathfrak{S}\big(\mathfrak{s}(x)O\boldsymbol{\xi}\big)_{\mathfrak{s}(x)}(F\circ R_O) - \lambda\Big(A\big(\mathfrak{s}(x)O\boldsymbol{\xi}\big)\Big)_{\mathfrak{s}(x)}(F\circ R_O)\\ &= \big(\mathfrak{s}(x)O\boldsymbol{\xi}\big)\tilde{F}(\,\cdot\,,O) - \sum_{i=1}^d (O\boldsymbol{\xi})^i\frac{d}{ds}\tilde{F}\left(x, e^{sa_i(x)}O\right)\Big|_{s=0},\end{aligned}$$

where $\tilde{F}(x,O) \equiv F\big(R_O\mathfrak{s}(x)\big)$ on $U\times O(\mathbb{R}^d)$ and (cf. (7.13))

$$a_i(x) \equiv \Big(\Big(\big[\Gamma^k_{i,j}(\mathbf{e})\big](x)\Big)\Big)_{1\le j,k\le d} \quad \text{for } 1 \le i \le d \text{ and } x \in U.$$

Hence, the $\lambda_{\mathcal{O}(M)}$-integral of $\mathfrak{E}(\boldsymbol{\xi})F$ over $\pi^{-1}(U)$ is equal to

$$\begin{aligned}&\int_{O(\mathbb{R}^d)}\left(\int_U \big(\mathfrak{s}(x)O\boldsymbol{\xi}\big)\tilde{F}(\,\cdot\,,O)\,\lambda_M(dx)\right)\lambda_{O(\mathbb{R}^d)}(dO)\\ &\qquad - \sum_{i=1}^d \frac{d}{ds}\int_U\left(\int_{O(\mathbb{R}^d)}(O\boldsymbol{\xi})^i\tilde{F}\left(x,e^{sa_i(x)}O\right)\,\lambda_{O(\mathbb{R}^d)}(dO)\right)\lambda_M(dx)\bigg|_{s=0}\\ &= \iint\limits_{U\times O(\mathbb{R}^d)}\left(-\big[\operatorname{div}^M(\mathfrak{s}O\boldsymbol{\xi})\big](x) + \sum_{i=1}^d\big(a_i(x)O\boldsymbol{\xi}\big)^i\right)\tilde{F}(x,O)\,\lambda_M(dx)\lambda_{O(\mathbb{R}^d)}(dO).\end{aligned}$$

Finally, by (7.19),

$$\big[\operatorname{div}^M(\mathfrak{s}O\boldsymbol{\xi})\big](x) = \sum_{i,j=1}^d (O\boldsymbol{\xi})^j\big\langle\nabla^M_{(E_i)_x}E_j,(E_i)_x\big\rangle = \sum_{i=1}^d\big(a_i(x)O\boldsymbol{\xi}\big)^i.$$

8.2 Brownian Motion on M via Projection from $\mathcal{O}(M)$

Throughout what follows, we will assume that M is a connected, d-dimensional, separable Riemannian manifold which is complete,[6] and we will be using (8.29) as the basis for a construction of the Brownian motion on M. That is, we will actually construct the process on $\mathcal{O}(M)$ and will then use the fiber map π to project it down to M.

8.2.1. The Basic Construction. Given $\mathfrak{f} \in \mathcal{O}(M)$ and $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$, determine[7] $\mathfrak{p}(\,\cdot\,, \mathfrak{f}, \mathbf{w}_n) \in \mathcal{P}(\mathcal{O}(M))$ by (8.27) after replacing $\mathbf{w}$ with (cf. (3.17)) $\mathbf{w}_n$.

8.33 THEOREM. *Given $x \in M$ and $\mathfrak{f} \in \pi^{-1}(x)$, the martingale problem for $\frac{1}{2}\Delta_M$ starting at x is well-posed if and only if the martingale for $\frac{1}{2}\Delta_B$ starting at $\mathfrak{f}$ is well-posed, in which case*

$$\mathbb{P}_x^M \equiv \mathbb{P}_x^{\frac{1}{2}\Delta_M} = \pi_* \mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)}, \quad \text{where } \mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)} \equiv \mathbb{P}_{\mathfrak{f}}^{\frac{1}{2}\Delta_B}.$$

Moreover, if $\mathbb{P}_x^M$ exists, then, for each $\mathfrak{f} \in \pi^{-1}(x)$, $\mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)}$ is the $\mu_{\mathbb{R}^d}$-distribution of a measurable map $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto \mathfrak{p}(\,\cdot\,, \mathfrak{f}, \mathbf{w}) \in \mathcal{P}(\mathcal{O}(M))$ to which the sequence $\{\mathfrak{p}(\,\cdot\,, \mathfrak{f}, \mathbf{w}_n)\}_0^\infty$ converges in $\mathcal{P}(\mathcal{O}(M))$ for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$. Finally, if the martingale problem for $\frac{1}{2}\Delta_M$ is well-posed on M, then the martingale problem for $\frac{1}{2}\Delta_B$ is well-posed on $\mathcal{O}(M)$, and there is a measurable $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto \mathfrak{p}(\,\cdot\,, *, \mathbf{w}) \in C^{0,\infty}([0,\infty) \times \mathcal{O}(M); \mathcal{O}(M))$ to which $\{\mathfrak{p}(\,\cdot\,, *, \mathbf{w}_n)\}_0^\infty$ converges in $C^{0,\infty}([0,\infty) \times \mathcal{O}(M); \mathcal{O}(M))$ for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$.

PROOF: In view of (8.29), the only aspect of this statement which is not covered by Theorem 7.2 is the very first one. On the other hand, the only obstruction to the existence of $\mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)}$ is the possibility of explosion. Hence, since the only way for a path $\mathfrak{p}$ on $\mathcal{O}(M)$ to explode is for the path $\pi \circ \mathfrak{p}$ to explode from M, there is nothing more to say. □

Assuming that the martingale problem for $\frac{1}{2}\Delta_M$ is well-posed on M, define the semigroup $\{\mathbf{P}_t^{\mathcal{O}(M)} : t \geq 0\}$ on $C_b(\mathcal{O}(M); \mathbb{R})$ by

$$[\mathbf{P}_t^{\mathcal{O}(M)} F](\mathfrak{f}) = \mathbb{E}^{\mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)}}\Big[F(\mathfrak{p}(t))\Big]. \tag{8.34}$$

Then, just as in Theorem 3.58, (8.32) implies that

$$\int_{\mathcal{O}(M)} G\, \mathbf{P}_t^{\mathcal{O}(M)} F \, d\lambda_{\mathcal{O}(M)} = \int_{\mathcal{O}(M)} F\, \mathbf{P}_t^{\mathcal{O}(M)} G \, d\lambda_{\mathcal{O}(M)}, \quad t \in (0,\infty), \tag{8.35}$$

---

[6] It has been pointed out to me by H. Sussman that general principals of point-set topology make it unnecessary to add "separability," it is automatic.

[7] Because M is complete, existence presents no problem.

for all $F, G \in C\big(\mathcal{O}(M);[0,\infty)\big)$. In particular, since $[\mathbf{P}_t^M f]\circ\pi = \mathbf{P}_t^{\mathcal{O}(M)}(f\circ\pi)$ and $\lambda_M = \pi_*\lambda_{\mathcal{O}(M)}$,

$$\int_M g\,\mathbf{P}_t^M f\,d\lambda_M = \int_M f\,\mathbf{P}_t^M g\,d\lambda_M, \quad t\in(0,\infty), \tag{8.36}$$

for all $f,g\in C\big(M;[0,\infty)\big)$.

Remark: It may be worth noting that the preceding construction of Brownian motion on M lends itself nicely to a Feynman-type (cf. §1.3) picture. Namely, let x_0 (alias the origin) be a fixed reference point in M, and, by analogy to $H^1(\mathbb{R}^d)$, define $H^1(M)$ to be the set of all absolutely continuous $p\in\mathcal{P}(M)$ with properties that $p(0)=x_0$ and $t\rightsquigarrow\|\dot{p}\|$ is square integrable. Next, choose and fix a reference frame $\mathfrak{f}_0\in\pi^{-1}x_0$, and define the *development map* $p\in H^1(M)\longrightarrow \mathbf{h}_p\in H^1(\mathbb{R}^d)$ so that

$$\mathfrak{f}_0\mathbf{h}_p(t) = \int_0^t \mathcal{T}_{p\restriction[0,\tau]}^{-1}\dot{p}(\tau)\,d\tau, \quad t\in[0,\infty).$$

Next, notice that the (cf. (8.24)) rolling map provides an inverse for the development map. More precisely, if $\mathcal{R}: H^1(\mathbb{R}^d)\longrightarrow H^1(M)$ is defined so that (cf. (8.27)) $p(\,\cdot\,)=\pi\circ\mathfrak{p}(\,\cdot\,,\mathfrak{f}_0,\mathbf{w})$, then $p=\mathcal{R}\mathbf{h}$ if and only if $\mathrm{h}=\mathbf{h}_p$. In addition,

$$\|p\|_{H^1(M)} \equiv \left(\int_0^\infty \|\dot{p}(t)\|^2\,dt\right)^{\frac{1}{2}} = \|\mathbf{h}_p\|_{H^1(\mathbb{R}^d)}.$$

Thus, if we were to pretend that $\mu_{\mathbb{R}^d}$ lives on $H^1(\mathbb{R}^d)$ and is given by Feynman's picture in (1.20), then the construction here of $\mathbb{P}_{x_0}^M$ would say that

$$\frac{d\mathbb{P}_{x_0}^M}{d\lambda_{H^1(M)}} = \frac{1}{Z_{H^1(M)}}\exp\left(-\frac{\|p\|_{H^1(M)}^2}{2}\right),$$

where $\lambda_{H^1(M)} = \mathcal{R}_*\lambda_{H^1(\mathbb{R}^d)}$ and $Z_{H^1(M)} = Z_{H^1(\mathbb{R}^d)}$.

8.2.2. Parallel Transport along Brownian Paths. Among the many advantages to the construction just given is the fact that it produces Brownian paths which are already equipped with a well-defined notion of parallel. To be precise, when $p\in\mathcal{P}(M)$ is smooth with $p(0)=x$ and $\mathfrak{p}\in\mathcal{P}\big(\mathcal{O}(M)\big)$ is the horizontal lift of p with $\mathfrak{p}(0)=\mathfrak{f}\in\pi^{-1}(x)$, then

$$\mathcal{T}_{p\restriction[0,t]}X_x = \mathfrak{p}(t)\big(\mathfrak{f}^{-1}(X_x)\big).$$

Now suppose that the martingale problem for $\frac{1}{2}\Delta_M$ starting at x is well-posed, and let $\mathfrak{f}\in\pi^{-1}(x)$ be given. Then, since $\mathfrak{p}(\,\cdot\,,\mathfrak{f},\mathbf{w}_n)$ is the horizontal lift of

$p(\cdot,\mathfrak{f},\mathbf{w}_n) \equiv \pi \circ \mathfrak{p}(\cdot,\mathfrak{f},\mathbf{w}_n)$ and, $\mu_{\mathbb{R}^d}$-almost surely, $\mathfrak{p}(\cdot,\mathfrak{f},\mathbf{w}_n) \longrightarrow \mathfrak{p}(\cdot,\mathfrak{f},\mathbf{w})$, we can think of $\mathfrak{p}(\cdot,\mathfrak{f},\mathbf{w})$ as being the horizontal lift, starting at $\mathfrak{f}$, of the path

$$p(\cdot,\mathfrak{f},\mathbf{w}) \equiv \pi \circ \mathfrak{p}(\cdot,\mathfrak{f},\mathbf{w}). \tag{8.37}$$

Equivalently, since $\mathbb{P}_x^M$ and $\mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)}$ are, respectively, the $\mu_{\mathbb{R}^d}$-distributions of $\mathbf{w}\rightsquigarrow p(\cdot,\mathfrak{f},\mathbf{w})$ and $\mathbf{w}\rightsquigarrow \mathfrak{p}(\cdot,\mathfrak{f},\mathbf{w})$, we are saying that $\mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)}$ is the distribution of horizontal lifts, starting at $\mathfrak{f}$, of Brownian paths starting at x. In particular, if $X_x \in T_xM$, then we will think of the distribution of

$$\mathfrak{p} \in \mathcal{P}\big(\mathcal{O}(M)\big) \longmapsto \mathfrak{p}(t)\big(\mathfrak{f}^{-1}X_x\big) \in T_{\pi\circ\mathfrak{p}(t)}M \quad \text{under } \mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)}$$

as the $\mathbb{P}_x^M$-distribution of

$$p \in \mathcal{P}(M) \longmapsto \mathcal{T}_{\pi\circ\mathfrak{p}\restriction[0,t]}X_x \in T_{p(t)}M.$$

For the sake of the reader who has noticed that there seems to be some ambiguity about the convention just suggested, we will conclude this discussion with an important remark about the relation between the $\mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)}$'s as $\mathfrak{f}$ runs over $\pi^{-1}(x)$. Namely, assuming that $\mathbb{P}_x^M$ exists at all,

$$\mathbb{P}_{R_O\mathfrak{f}}^{\mathcal{O}(M)} = (R_O)_*\mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)}, \quad O \in O(\mathbb{R}^d). \tag{8.38}$$

Indeed, because $\Delta_B(F \circ R_O) = [\Delta_B F] \circ R_O$, it is easy to check that the distribution of $\mathfrak{p}\rightsquigarrow R_O\mathfrak{p}$ under $\mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)}$ solves the martingale problem for $\frac{1}{2}\Delta_B$ starting at $R_O\mathfrak{p}$, and therefore (8.38) is an immediate consequence of the fact that there is only one solution to the martingale problem for $\frac{1}{2}\Delta_B$ starting at $R_O\mathfrak{f}$. Alternatively, and perhaps more interesting, is the following derivation of (8.38). To wit, use (8.14) and (8.27) to see that

$$R_O \circ \mathfrak{p}(\cdot,\mathfrak{f},\mathbf{w}_n) = \mathfrak{p}(\cdot,R_O\mathfrak{f},O^\top\mathbf{w}_n) \tag{8.39}$$

for each n. Hence, since the $\mu_{\mathbb{R}^d}$ is invariant under the transformation $\mathbf{w}\rightsquigarrow O^\top\mathbf{w}$, $\mathbf{w}\rightsquigarrow R_O \circ \mathfrak{p}(\cdot,\mathfrak{f},\mathbf{w})$ has the same $\mu_{\mathbb{R}^d}$-distribution as $\mathbf{w}\rightsquigarrow\mathfrak{p}(\cdot,R_O\mathfrak{f},\mathbf{w})$.

8.2.3. Measurability Considerations. [8] Having raised such questions in § 4.3 and again in § 6.3.3, it is only reasonable that we inquire into measurability questions connected with the present construction. In particular, one should ask whether, under $\mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)}$, the "horizontally lifted" path $\mathfrak{p}$ is a measurable function of the path $p \equiv \pi \circ \mathfrak{p}$. To be more precise, what we would like is to find a sequence of adapted maps $\Psi_n : [0,\infty) \times \mathcal{P}(M) \longrightarrow \mathcal{P}\big(\mathcal{O}(M)\big)$ such that $\Psi_n(\cdot,\pi\circ\mathfrak{p}) \longrightarrow \mathfrak{p}$ in $\mathcal{P}\big(\mathcal{O}(M)\big)$ for $\mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)}$-almost every $\mathfrak{p} \in \mathcal{P}\big(\mathcal{O}(M)\big)$.

[8] Like earlier excursions of this nature, this one can be safely skipped.

At first glance, one might say that we already have such a procedure. Namely, because $\mathfrak{p}(\,\cdot\,,\mathfrak{f},\mathbf{w}_n)$ is the horizontal lift of the piecewise smooth path $p(\,\cdot\,,\mathfrak{f},\mathbf{w}_n)$ and $\mathfrak{p}(\,\cdot\,,\mathfrak{f},\mathbf{w}_n) \longrightarrow \mathfrak{p}(\,\cdot\,,\mathfrak{f},\mathbf{w})$ for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$, one would like to take Ψ_n so that $\Psi_n\big(t,\pi\circ\mathfrak{p}(\,\cdot\,,\mathfrak{f},\mathbf{w})\big) = \mathfrak{p}(t,\mathfrak{f},\mathbf{w}_n)$. However, justification of this definition is tantamount to showing that, at least up to a $\mu_{\mathbb{R}^d}$-null set, $\mathbf{w}$ is a measurable function of $p(\,\cdot\,,\mathfrak{f},\mathbf{w})$. Hence, what we really ought to be doing is trying to construct a measurable function for recovering $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$ from the corresponding path $p(\,\cdot\,,\mathfrak{f},\mathbf{w})$.

The following lemma is the essential ingredient which we will use to make this construction.

8.40 LEMMA. *Let M and N be a pair of separable differentiable manifolds, and use π_M and π_N to denote the natural projection maps from $M\times N$ onto M and N, respectively. Further, assume that $(Z_0,\dots,Z_r)$ are vector fields on $M\times N$ for which there exists a smooth map $(x,y)\in M\times N \longmapsto \Lambda_{(x,y)}\in \operatorname{Hom}(T_xM;T_yN)$ such that*

$$(\pi_N)_*(Z_k)_{(x,y)} = \Lambda_{(x,y)}\big((\pi_M)_*(Z_k)_{(x,y)}\big) \quad \text{for } 1\le k\le r.$$

If $\mathbb{Q}_z$ solves the martingale problem for $Z_0+\frac12\sum_{k=1}^r Z_k^2$ starting at some point $z=(x,y)\in M\times N$, then there exists an adapted map $\Psi:[0,\infty)\times\mathcal{P}(M)\longrightarrow N$ such that $\Psi(\,\cdot\,,p)$ is continuous for $(\Pi_M)_*\mathbb{Q}_z$-almost every $p\in\mathcal{P}(M)$ and

$$\mathbb{Q}_z\Big(\big\{q\in\mathcal{P}(M\times N)\,;\,\Pi_N\circ q(t)=\Psi\big(t,\Pi_M\circ q\big) \text{ for all } t\in[0,\infty)\big\}\Big)=1.$$

PROOF: In order to avoid unnecessary complications, we may and will assume that M is a closed, embedded submanifold of $\mathbb{R}^L$ for some $L\ge 1$ and that the Z_k's all have compact support. We extend $\Lambda_{(x,y)}$ to become an element of $\operatorname{Hom}(\mathbb{R}^L;T_yN)$ by setting $\Lambda_{(x,y)}(\boldsymbol{\eta})=\Lambda_{(x,y)}(\Pi_x^M\boldsymbol{\eta})$, where $\Pi_x^M:\mathbb{R}^L\longrightarrow T_xM$ denotes orthogonal projection.

Given any $p\in\mathcal{P}(M)$, define $p_n\in\mathcal{P}(\mathbb{R}^L)$ for $n\in\mathbb{N}$ so that, for each $m\in\mathbb{N}$ and $t\in\big[m2^{-n},(m+1)2^{-n}\big]$,

$$p_n(t)=\big((m+1)-2^nt\big)p(m2^{-n})+\big(2^nt-m\big)p\big((m+1)2^{-n}\big);$$

and determine $\Psi_n(\,\cdot\,,p)\in\mathcal{P}(N)$ from

$$\dot\Psi_n(t,p)=\Lambda_{(p(t),\Psi_n(t,p))}\big(\dot p_n(t)\big)+(\pi_N)_*(Z_0)_{(p(t),\Psi_n(t,p))} \quad \text{with } \Psi_n(0,p)=y.$$

Then, by essentially the same line of reasoning as we used in § 2.2, one can show that there is a progressively measurable $\Psi:[0,\infty)\times\mathcal{P}(M)\longrightarrow N$ such that

$$\Psi_n(\,\cdot\,,\Pi_M\circ q)\longrightarrow\Psi(\,\cdot\,,\pi_M\circ q) \quad \text{in } \mathcal{P}(N) \text{ for } \mathbb{Q}_z\text{-almost every } q\in\mathcal{P}(M\times N).$$

Moreover, the $\mathbb{Q}_z$-distribution of $q\in\mathcal{P}(M)\longmapsto\big(\Pi_Mq,\Psi(\,\cdot\,,\Pi_M\circ q)\big)\in\mathcal{P}(M\times N)$ solves the martingale problem for $Z_0+\frac12\sum_{k=1}^r Z_k^2$ starting at z. Thus, because there is only one such solution, we can conclude that $\Pi_N q = \Psi(\,\cdot\,,\Pi_M \circ q)$ for $\mathbb{Q}_z$-almost every $q \in \mathcal{P}(M \times N)$. $\square$

In order to apply this to the question at hand, we will begin with the case in which $\mathcal{O}(M)$ is *trivial* in the sense that there exists a smooth, global section

$$x \in M \longmapsto \mathfrak{s}(x) = \bigl(x, \mathbf{e}(x)\bigr) = \Bigl(x, \bigl((E_1)_x, \dots, (E_d)_x\bigr)\Bigr) \in \mathcal{O}(M),$$

in which case we can identify $\mathcal{O}(M)$ with $M \times O(\mathbb{R}^d)$ via the correspondence $M \times O(\mathbb{R}^d) \ni (x, O) \longleftrightarrow R_O\mathfrak{s}(x) \in \mathcal{O}(M)$. Now take $N = O(\mathbb{R}^d) \times \mathbb{R}^d$, choose an orthonormal basis $\mathfrak{e} = (\mathbf{e}_1, \dots, \mathbf{e}_d)$ in $\mathbb{R}^{d'}$, and define the vector fields Z_k, $1 \le k \le d$, on $M \times N$ so that, for $\varphi \in C^\infty(M \times N; \mathbb{R})$,

$$\begin{aligned}(Z_k)_{(x,(O,\boldsymbol{\xi}))}\varphi = {} & \bigl(\mathfrak{s}(x)O\mathbf{e}_k\bigr)\varphi\bigl(\,\cdot\,,(O,\boldsymbol{\xi})\bigr) \\ & - \rho\Bigl(A\bigl(\mathfrak{s}(x)O\mathbf{e}_k\bigr)\Bigr)_O \varphi\bigl(x, (\,\cdot\,,\boldsymbol{\xi})\bigr) + \partial_k^{\mathfrak{e}}\varphi\bigl(x,(O,\,\cdot\,)\bigr),\end{aligned}$$

where $A(X_x)$ is defined as in (8.9) and, for $a \in \mathfrak{o}(\mathbb{R}^d)$, $\rho(a)$ is the right invariant vector field on $O(\mathbb{R}^d)$ defined by $\rho(a)_O = \frac{d}{ds}e^{sa}O\big|_{s=0}$. Then, if $F : M \times N \longrightarrow \mathcal{O}(M) \times \mathbb{R}^d$ is the diffeomorphism defined by $F\bigl(x,(O,\boldsymbol{\xi})\bigr) = \bigl(R_O\mathfrak{s}(x), \boldsymbol{\xi}\bigr) \in \mathcal{O}(M) \times \mathbb{R}^d$, we have (cf. (8.13))

$$F_*(Z_k)_{(x,(O,\boldsymbol{\xi}))}\varphi = \mathfrak{E}(\mathbf{e}_k)_{R_O\mathfrak{s}(x)}\varphi(\,\cdot\,,\boldsymbol{\xi}) + (\partial_k^{\mathfrak{e}})_{\boldsymbol{\xi}}\varphi\bigl(R_O\mathfrak{s}(x),\,\cdot\,\bigr).$$

Hence, $\mathbb{Q}_{(x,(O,\boldsymbol{\xi}))}$ on $\mathcal{P}(M \times N)$ solves the martingale problem for $\frac{1}{2}\sum_{k=1}^d Z_k^2$ starting at $\bigl(x,(O,\mathbf{0})\bigr)$ if and only if the $\mathbb{Q}_{(x,(O,\mathbf{0}))}$-distribution of $q \in \mathcal{P}(M \times N) \longmapsto F \circ q \in \mathcal{P}\bigl(\mathcal{O}(M) \times \mathbb{R}^d\bigr)$ is a solution to the martingale problem for $\frac{1}{2}\sum_{k=1}^d \bigl(\mathfrak{E}(\mathbf{e}_k) + \partial_k^{\mathfrak{e}}\bigr)^2$ starting at $\bigl(R_O\mathfrak{s}(x), \mathbf{0}\bigr)$. But the $\mu_{\mathbb{R}^d}$-distribution of

$$\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto \bigl(\mathfrak{p}(\,\cdot\,, R_O\mathfrak{s}(x), \mathbf{w}), \mathbf{w}\bigr) \in \mathcal{P}\bigl(\mathcal{O}(M) \times \mathbb{R}^d\bigr)$$

also solves this martingale starting at the same place. Hence, if there exists a measurable map $\Psi : [0,\infty) \times \mathcal{P}(M) \longrightarrow N$ such that $\pi_N \circ q = \Psi(\,\cdot\,, \pi_M \circ q)$ for $Q_{(x,(O,\mathbf{0}))}$-almost every $q \in \mathcal{P}(M \times N)$, then

$$\mathbf{w} = \pi_{\mathbb{R}^d} \circ \Psi\bigl(\,\cdot\,, p(\,\cdot\,,\mathfrak{f},\mathbf{w})\bigr) \quad \text{for } \mu_{\mathbb{R}^d}\text{-almost all } \mathbf{w} \in \mathfrak{W}(\mathbb{R}^d).$$

Finally, because $(\pi_N)_*(Z_k)_{(x,(O,\boldsymbol{\xi}))}$ can be expressed as

$$\sum_{\ell=1}^d \bigl\langle \mathfrak{s}(x)O\mathbf{e}_\ell, (\pi_M)_*(Z_k)_{(x,(O,\boldsymbol{\xi}))}\bigr\rangle \Bigl(-\rho\bigl(A(\mathfrak{s}(x)\mathbf{e}_\ell)\bigr) + (\partial_\ell^{\mathfrak{e}})_{\boldsymbol{\xi}}\Bigr),$$

Lemma 8.40 guarantees the existence of Ψ.

When $\mathcal{O}(M)$ is not trivial, the preceding argument has to be localized. That is, one has to use the fact that each $x \in M$ admits a neighborhood V for which

$\pi^{-1}(V)$ is trivial. Thus one can choose a countable cover $\{U_m\}$ of M by relatively compact open sets U_m such that, for each m, $\bar{U}_m \subseteq V_m$ where V_m is an open set on which there exists a smooth section $x \in V_m \longmapsto \mathfrak{s}_m(x) \in \mathcal{O}(M)$. Now, define $m(x) = \min\{m : x \in U_m\}$, and, for $p \in \mathcal{P}(M)$, define $\{\sigma_n(p)\}_0^\infty$ so that $\sigma_0(p) = 0$ and, for $n \geq 1$,

$$\sigma_n(p) = \begin{cases} \infty & \text{if } \sigma_{n-1}(p) = \infty \\ \inf\{t \geq \sigma_{n-1}(p) : p(t) \notin U_{m(p(\sigma_{n-1}))}\} & \text{if } \sigma_{n-1}(p) < \infty. \end{cases}$$

Setting $\tau_n(\mathbf{w}) = \sigma_{n-1} \circ p(\,\cdot\,, \mathfrak{f}, \mathbf{w})$ and using the flow property (cf. § 2.5.1):

$$\mathfrak{p}\big(\tau_{n-1}(\mathbf{w}) + t, \mathfrak{f}, \mathbf{w}\big) = \mathfrak{p}\Big(t, \mathfrak{p}\big(\tau_{n-1}(\mathbf{w}), \mathfrak{f}, \mathbf{w}\big), \delta_{\tau_{n-1}(\mathbf{w})}\mathbf{w}\Big) \quad (\text{a.s.}, \mu_{\mathbb{R}^d}),$$

one can apply the preceding to build an inductive procedure to see that, for each $n \in \mathbb{N}$, $\mathbf{w} \restriction \big[0, \sigma_n(p(\,\cdot\,, \mathfrak{f}, \mathbf{w})\big)$ is a progressively measurable function of $p(\,\cdot\,, \mathfrak{f}, \mathbf{w}) \restriction \big[0, \sigma_n(p(\,\cdot\,, \mathfrak{f}, \mathbf{w})\big)$. Hence, since $\sigma_n(p) \nearrow \infty$, we have now proved the following important fact.

8.41 THEOREM. *Let $\mathfrak{f} \in \mathcal{O}(M)$ and assume that $\mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)}$ exists. Then there is an adapted map $\Psi_{\mathfrak{f}} : [0, \infty) \times \mathcal{P}(M) \longrightarrow \mathfrak{W}(\mathbb{R}^d)$ such that $\Psi_{\mathfrak{f}}(\,\cdot\,, \pi \circ \mathfrak{p})$ is continuous for $\mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)}$-almost every $\mathfrak{p} \in \mathcal{P}\big(\mathcal{O}(M)\big)$ and $\mathbf{w} = \Psi_{\mathfrak{f}}\big(\,\cdot\,, p(\,*\,, \mathfrak{f}, \mathbf{w})\big)$ for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$. In particular, up to a $\mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)}$-null set, $\mathfrak{p}(\,\cdot\,, \mathfrak{f}, \mathbf{w}) \restriction [0, T]$ is a measurable function of $p(\,\cdot\,, \mathfrak{f}, \mathbf{w}) \restriction [0, T]$ for each $T \in [0, \infty)$.*

8.3 Curvature Considerations and an Explosion Criterion

In this section we will reproduce the results of § 5.2 in the present context, and the first step is to see what the Riemann curvature looks like in terms of the structure of $\mathcal{O}(M)$.

8.3.1. Cartan's Structural Equations. We begin by recalling the *exterior differential* operator "d" of differential geometry. Namely, given an n-form η on a manifold, $d\eta$ is the $(n+1)$-form such that $d\eta(X) = X\eta$ if $n = 0$ (and therefore η is a function),

$$(8.42) \qquad d\eta(X_1, X_2) = X_1\eta(X_2) - X_2\eta(X_1) - \eta([X_1, X_2]) \quad \text{if } n = 1,$$

and, for general $n \geq 1$, if $\mathbf{X} = (X_1, \dots, X_{n+1})$, then

$$(8.43) \quad d\eta(\mathbf{X}) = \sum_{m=1}^{n+1} (-1)^{m+1} X_m \eta(\mathbf{X}^{(m)}) + \sum_{1 \leq \ell < m \leq n+1} (-1)^{\ell+m} \eta(\mathbf{X}^{(\ell,m)}),$$

where $\mathbf{X}^{(m)} \equiv (X_1, \dots, \hat{X}_m, \dots, X_{n+1})$,

$$\mathbf{X}^{(\ell,m)} \equiv \big([X_\ell, X_m], X_1, \dots, \hat{X}_\ell, \dots, \hat{X}_m, \dots, X_{n+1}\big),$$

and the superscripts "ˆ" are used to indicate that the corresponding vector does not appear but that the order of the remaining vectors has not been altered. Perhaps the most fundamental fact about this operation is that $d\eta$ is indeed an $(n+1)$-form. That is, in spite of presentation in terms of vector fields, the value of $d\eta(X_1,\dots,X_{n+1})$ at a point x depends only on $\big((X_1)_x,\dots,(X_{n+1})_x\big)$ and not on the vector fields X_k anywhere else. Equivalently,

$$d\eta\big(X_1,\dots,fX_m,\dots,X_{n+1}\big) = f\,d\eta\big(X_1,\dots,X_m,\dots,X_{n+1}\big)$$

for any $1 \le m \le n$ and smooth function f. Thus, we can, and will, unambiguously write $d\eta\big((X_1)_x,\dots,(X_{n+1})_x\big)$ for the value of $d\eta(X_1,\dots,X_{n+1})$ at x. A second fundamental fact is that two applications of "d" is lethal in the sense that $d^2\eta = 0$ for any η. A third comment is that, although n-forms are traditionally thought of as real or complex valued, they can take values in any linear space, and all the properties of "d" transfer to this general setting.[9]

We now want to apply "d" to the $\mathbb{R}^d$-valued solder 1-form ϕ (cf. (8.16)) and the $o(\mathbb{R}^d)$-valued connection 1-form ω (cf. (8.17)). In view of the fact that $d\phi(\mathfrak{X}_\mathfrak{f},\mathfrak{Y}_\mathfrak{f})$ and $d\omega(\mathfrak{X}_\mathfrak{f},\mathfrak{Y}_\mathfrak{f})$ depend only on the values of $\mathfrak{X}$ and $\mathfrak{Y}$ at $\mathfrak{f}$ and are bilinear in those values, we need only consider the cases:

(i) $\mathfrak{X} = \mathfrak{E}(\boldsymbol{\xi})$ and $\mathfrak{Y} = \mathfrak{E}(\boldsymbol{\eta})$ for some $\boldsymbol{\xi},\boldsymbol{\eta} \in \mathbb{R}^d$
(ii) $\mathfrak{X} = \mathfrak{E}(\boldsymbol{\xi})$ and $\mathfrak{Y} = \lambda(a)$ for some $\boldsymbol{\xi} \in \mathbb{R}^d$ and $a \in o(\mathbb{R}^d)$
(iii) $\mathfrak{X} = \lambda(a)$ and $\mathfrak{Y} = \lambda(b)$ for some $a,b \in o(\mathbb{R}^d)$.

Since $\phi\big(\mathfrak{E}(\boldsymbol{\xi})\big) \equiv \boldsymbol{\xi}$ and $\phi\big(\mathfrak{E}(\boldsymbol{\eta})\big) \equiv \boldsymbol{\eta}$, while $\omega\big(\mathfrak{E}(\boldsymbol{\xi})\big) \equiv 0 \equiv \omega\big(\mathfrak{E}(\boldsymbol{\eta})\big)$, (8.42) says that

$$(*)\qquad \begin{aligned} d\phi\big(\mathfrak{E}(\boldsymbol{\xi}),\mathfrak{E}(\boldsymbol{\eta})\big) &= -\phi\Big(\big[\mathfrak{E}(\boldsymbol{\xi}),\mathfrak{E}(\boldsymbol{\eta})\big]\Big)\\ d\omega\big(\mathfrak{E}(\boldsymbol{\xi}),\mathfrak{E}(\boldsymbol{\eta})\big) &= -\omega\Big(\big[\mathfrak{E}(\boldsymbol{\xi}),\mathfrak{E}(\boldsymbol{\eta})\big]\Big). \end{aligned}$$

To go further, we must observe that, because of (8.30) and (7.22), $\pi_*\big[\mathfrak{E}(\boldsymbol{\xi}),\mathfrak{E}(\boldsymbol{\eta})\big] \equiv 0$. Equivalently, there exists an $o(\mathbb{R}^d)$-valued 2-form $\mathfrak{f}\rightsquigarrow\Omega_\mathfrak{f}$ such that

$$(8.44)\qquad \big[\mathfrak{E}(\boldsymbol{\xi}),\mathfrak{E}(\boldsymbol{\eta})\big] = -\lambda\Big(\Omega(\boldsymbol{\xi},\boldsymbol{\eta})\Big),$$

which, for reasons which will become obvious shortly, is called the *curvature 2-form*. Notice that, by the second part of (*), we have

$$(8.45)\qquad \mathfrak{X}_\mathfrak{f},\mathfrak{Y}_\mathfrak{f} \in H_\mathfrak{f}\mathcal{O}(M) \implies d\omega(\mathfrak{X}_\mathfrak{f},\mathfrak{Y}_\mathfrak{f}) = \Omega_\mathfrak{f}\Big(\phi(\mathfrak{X}_\mathfrak{f}),\phi(\mathfrak{Y}_\mathfrak{f})\Big).$$

At the same time, because $\big[\mathfrak{E}(\boldsymbol{\xi}),\mathfrak{E}(\boldsymbol{\eta})\big]$ is vertical, the first part of (*) now says that

$$\mathfrak{X}_\mathfrak{f},\mathfrak{Y}_\mathfrak{f} \in H_\mathfrak{f}\mathcal{O}(M) \implies d\phi\Big(\mathfrak{X}_\mathfrak{f},\mathfrak{Y}_\mathfrak{f}\Big) = 0,$$

[9] A good source for the basic theory is [24].

We turn next to cases (ii) and (iii). But, since $\phi(\lambda(b)) \equiv \mathbf{0}$ while $\omega(\lambda(b)) \equiv b$, (8.42) together with (8.15) say that

$$d\phi(\mathfrak{E}(\boldsymbol{\xi}), \lambda(b)) = -b\boldsymbol{\xi} \quad \text{and} \quad d\omega(\mathfrak{E}(\boldsymbol{\xi}), \lambda(b)) = 0.$$

Similarly, by (8.5),

$$d\phi(\lambda(a), \lambda(b)) = \mathbf{0} \quad \text{and} \quad d\omega(\lambda(a), \lambda(b)) = -[a, b].$$

After combining these, we obtain Cartan's *first structural equation*

$$d\phi(\mathfrak{X}_{\mathfrak{f}}, \mathfrak{Y}_{\mathfrak{f}}) = -\omega \wedge \phi(\mathfrak{X}_{\mathfrak{f}}, \mathfrak{Y}_{\mathfrak{f}}) \tag{8.46}$$

and Cartan's *second structural equation*

$$d\omega(\mathfrak{X}_{\mathfrak{f}}, \mathfrak{Y}_{\mathfrak{f}}) = -\omega \wedge \omega(\mathfrak{X}_{\mathfrak{f}}, \mathfrak{Y}_{\mathfrak{f}}) + \Omega_{\mathfrak{f}}\Big(\phi(\mathfrak{X}_{\mathfrak{f}}), \phi(\mathfrak{Y}_{\mathfrak{f}})\Big), \tag{8.47}$$

where $\eta_1 \wedge \eta_2(X, Y) \equiv \eta_1(X)\eta_2(Y) - \eta_1(Y)\eta_2(X)$. In particular, if $(s,t) \rightsquigarrow \mathsf{p}(s,t)$ is a twice differentiable, two parameter map and if $\mathsf{p}'(s,t)$ and $\dot{\mathsf{p}}(s,t)$ denote, respectively, $\frac{\partial \mathsf{p}}{\partial s}(s,t)$ and $\frac{\partial \mathsf{p}}{\partial t}(s,t)$, then (because ∂_s commutes with ∂_t)

$$\begin{aligned} &\frac{d}{ds}\phi(\dot{\mathsf{p}}(s,t)) - \frac{d}{dt}\phi(\mathsf{p}'(s,t)) \\ &\qquad = -\omega(\mathsf{p}'(s,t))\phi(\dot{\mathsf{p}}(s,t)) + \omega(\dot{\mathsf{p}}(s,t))\phi(\mathsf{p}'(s,t)) \end{aligned} \tag{8.48}$$

and

$$\begin{aligned} &\frac{d}{ds}\omega(\dot{\mathsf{p}}(s,t)) - \frac{d}{dt}\omega(\mathsf{p}'(s,t)) \\ &\qquad = -\Big[\omega(\mathsf{p}'(s,t)), \omega(\dot{\mathsf{p}}(s,t))\Big] + \Omega_{\mathsf{p}(s,t)}\Big(\phi(\mathsf{p}'(s,t)), \phi(\dot{\mathsf{p}}(s,t))\Big) \end{aligned} \tag{8.49}$$

Finally, there are two identities involving Ω which it is useful to have available. The first of these is the statement that the curvature 2-form transforms equivariantly under the action of $O(\mathbb{R}^d)$. That is,

$$\Omega_{R_O\mathfrak{f}}(\boldsymbol{\xi}, \boldsymbol{\eta}) = O^\top \Omega_{\mathfrak{f}}(O\boldsymbol{\xi}, O\boldsymbol{\eta})O. \tag{8.50}$$

To see this, first use (8.14) to see that

$$\big[\mathfrak{E}(\boldsymbol{\xi}), \mathfrak{E}(\boldsymbol{\eta})\big]_{R_O\mathfrak{f}} = (R_O)_*\big[\mathfrak{E}(O\boldsymbol{\xi}), \mathfrak{E}(O\boldsymbol{\eta})\big]_{\mathfrak{f}},$$

and then apply the second part of (8.19) to get (8.50) from (cf. (8.44))

$$\Omega(\boldsymbol{\xi}, \boldsymbol{\eta}) = -\omega([\mathfrak{E}(\boldsymbol{\xi}), \mathfrak{E}(\boldsymbol{\eta})]). \tag{8.51}$$

The second identity is called the *second Bianci identity* and is a little more sophisticated. Namely, we given $\xi_1, \xi_2, \xi_3 \in \mathbb{R}^d$, we have

$$(8.52) \qquad \mathfrak{E}(\xi_1)\Omega(\xi_2,\xi_3) + \mathfrak{E}(\xi_2)\Omega(\xi_3,\xi_1) + \mathfrak{E}(\xi_3)\Omega(\xi_1,\xi_2) = 0.$$

The proof of (8.52) is an application of (8.47), the fact (cf. (8.44)) that the commutator of canonical vector fields is vertical, and the elementary fact that d^2 annihilates any form. More explicitly, we have (cf. (8.43) with $n = 2$):

$$\begin{aligned} 0 &= d^2\omega\big(\mathfrak{E}(\xi_1),\mathfrak{E}(\xi_2),\mathfrak{E}(\xi_3)\big) \\ &= \mathfrak{E}(\xi_1)d\omega\big(\mathfrak{E}(\xi_2),\mathfrak{E}(\xi_3)\big) - d\omega\big(\big[\mathfrak{E}(\xi_2),\mathfrak{E}(\xi_3)\big],\mathfrak{E}(\xi_1)\big) \\ &\qquad - \mathfrak{E}(\xi_2)d\omega\big(\mathfrak{E}(\xi_1),\mathfrak{E}(\xi_3)\big) + d\omega\big(\big[\mathfrak{E}(\xi_1),\mathfrak{E}(\xi_3)\big],\mathfrak{E}(\xi_2)\big) \\ &\qquad\quad + \mathfrak{E}(\xi_3)d\omega\big(\mathfrak{E}(\xi_1),\mathfrak{E}(\xi_2)\big) - d\omega\big(\big[\mathfrak{E}(\xi_1),\mathfrak{E}(\xi_2)\big],\mathfrak{E}(\xi_3)\big) \\ &= \mathfrak{E}(\xi_1)\Omega(\xi_2,\xi_3) - \mathfrak{E}(\xi_2)\Omega(\xi_1,\xi_3) + \mathfrak{E}(\xi_3)\Omega(\xi_1,\xi_2), \end{aligned}$$

which, by the skew symmetry of Ω, is equivalent to (8.52).

8.3.2. Riemann and Ricci Curvatures. With these preparations, we are now ready to see how the *Riemann curvature* tensor appears in $\mathcal{O}(M)$. Thus, once again (cf. (5.24)) we start with the definition

$$(8.53) \qquad R(X,Y)Z = \big[\nabla^M_X,\nabla^M_Y\big]Z - \nabla^M_{[X,Y]}Z$$

for vector fields X, Y, and Z on M. To check that, for each $x \in M$, $\big(R(X,Y)Z\big)_x$ is actually a trilinear function of $(X_x, Y_x, Z_x) \in (T_xM)^3$, take $\mathfrak{X}$ and $\mathfrak{Y}$ to be the horizontal lifts of the vector fields X and Y. Then, by (8.23),

$$\mathfrak{f}^{-1}\big[\nabla_X,\nabla_Y\big]_{\pi\mathfrak{f}}Z = \big[\mathfrak{X},\mathfrak{Y}\big]_{\mathfrak{f}}\Xi_Z.$$

At the same time, because $\pi_*[\mathfrak{X},\mathfrak{Y}]_{\mathfrak{f}} = [X,Y]_{\pi\mathfrak{f}}$ for all $\mathfrak{f} \in \mathcal{O}(M)$, another application of (8.23) shows that

$$\mathfrak{f}^{-1}\nabla^M_{[X,Y]_{\mathfrak{f}}}Z = [\mathfrak{X},\mathfrak{Y}]_{\mathfrak{f}}\Xi_Z + \omega\big([\mathfrak{X},\mathfrak{Y}]_{\mathfrak{f}}\big)\Xi_Z(\mathfrak{f}).$$

Hence, by (8.45) and (8.47), if $\mathfrak{f} \in \pi^{-1}(x)$, then

$$\begin{aligned} \mathfrak{f}^{-1}\Big(\big(R(X,Y)Z\big)_x\Big) &= -\omega([\mathfrak{X},\mathfrak{Y}]_{\mathfrak{f}})\mathfrak{f}^{-1}Z_x = d\omega(\mathfrak{X}_{\mathfrak{f}},\mathfrak{Y}_{\mathfrak{f}})\mathfrak{f}^{-1}Z_x \\ &= \Omega_{\mathfrak{f}}(\mathfrak{f}^{-1}X_x,\mathfrak{f}^{-1}Y_x)\mathfrak{f}^{-1}Z_x. \end{aligned}$$

Equivalently, we have now shown that, for $\mathfrak{f} \in \pi^{-1}(x)$,

$$(8.54) \qquad R(X_x,Y_x)Z_x \equiv \big(R(X,Y)Z\big)_x = \mathfrak{f}\Big(\Omega_{\mathfrak{f}}(\mathfrak{f}^{-1}X_x,\mathfrak{f}^{-1}Y_x)\mathfrak{f}^{-1}Z_x\Big),$$

which provides not only explicit evidence that the Riemann curvature at a point depends only on the vectors at that point but also justification for the name given to Ω.

Clearly, for each $x \in M$ and pair $(X_x, Y_x) \in T_xM$, we can think of $R(X_x, Y_x)$ as a linear transformation on T_xM. Moreover, $R(X_x, Y_x)$ satisfies the same symmetry relations as those established in § 5.2.3:

$$\text{(R1)} \qquad R(X_x, Y_x) = -R(Y_x, X_x)$$

$$\text{(R2)} \qquad R(X_x, Y_x)Z_x + R(Y_x, Z_x)X_x + R(Z_x, X_x)Y_x = 0$$

$$\text{(R3)} \qquad \langle R(X_x, Y_x)Z_x, W_x\rangle = -\langle R(X_x, Y_x)W_x, Z_x\rangle$$

$$\text{(R4)} \qquad \langle R(X_x, Y_x)Z_x, W_x\rangle = \langle R(Z_x, W_x)X_x, Y_x\rangle$$

Indeed, (R1) and (R3) are immediate from (8.54). To prove (R2), note that, by (8.44) and (8.15),

$$\Big[\mathfrak{E}(\boldsymbol{\zeta}), \big[\mathfrak{E}(\boldsymbol{\xi}), \mathfrak{E}(\boldsymbol{\eta})\big]\Big] = \mathfrak{E}\big(\Omega(\boldsymbol{\xi}, \boldsymbol{\eta})\boldsymbol{\zeta}\big),$$

and so (8.54) is equivalent to

$$R(X_x, Y_x)Z_x = \pi_*\Big(\Big[\mathfrak{E}(\mathfrak{f}^{-1}Z_x), \big[\mathfrak{E}(\mathfrak{f}^{-1}X_x), \mathfrak{E}(\mathfrak{f}^{-1}Y_x)\big]\Big]_{\mathfrak{f}}\Big) \tag{8.55}$$

for $\mathfrak{f} \in \pi^{-1}(U)$. Thus, (R2), which is the *first Bianci identity*, becomes an application of Jacobi's identity for the Lie algebra of vector fields on $\mathcal{O}(M)$. Finally, as we mentioned in the footnote about (R4) in § 5.2.3, (R4) is a consequence of (R1), (R2), and (R3).

For exactly the same reason as (5.26) held in the context there, it also holds here. Equivalently, if $(s,t) \rightsquigarrow \mathfrak{p}(s,t)$ and $(s,t) \rightsquigarrow \boldsymbol{\xi}(s,t)$ are smooth, two parameter maps into, respectively, $\mathcal{O}(M)$ and $\mathbb{R}^d$, then

$$\begin{aligned} &\frac{D^2\mathfrak{p}(s,t)\boldsymbol{\xi}(s,t)}{\partial s\partial t} - \frac{D^2\mathfrak{p}(s,t)\boldsymbol{\xi}(s,t)}{\partial t\partial s} \\ &\qquad = \mathfrak{p}(s,t)\Big(\Omega_{\mathfrak{p}(s,t)}\Big(\phi\big(\mathfrak{p}'(s,t)\big), \phi\big(\dot{\mathfrak{p}}(s,t)\big)\Big)\boldsymbol{\xi}(s,t)\Big). \end{aligned} \tag{8.56}$$

Given the Riemann curvature, the *Ricci curvature* is defined by the same prescription as we used in (5.27). That is, given a point $x \in M$, we define $\mathrm{Ric}_x \in \mathrm{Hom}(T_xM; T_xM)$ so that

$$\big\langle \mathrm{Ric}_x X_x, Y_x\big\rangle = \sum_{k=1}^{d} \big\langle R\big(X_x, (E_k)_x\big)(E_k)_x, Y_x\big\rangle, \tag{8.57}$$

for any orthonormal basis $\big((E_1)_x, \dots, (E_d)\big)$ in T_xM. Again, (R4) guarantees that Ric_x is symmetric. Notice that, for any choice of orthonormal basis $\mathfrak{e} = (\mathbf{e}_1, \dots, \mathbf{e}_d)$ in $\mathbb{R}^d$,

$$\begin{aligned} &\mathrm{Ric}_x X_x = \mathfrak{f}\mathfrak{R}_{\mathfrak{f}}\big(\mathfrak{f}^{-1}X_x\big), \quad \mathfrak{f} \in \pi^{-1}(x), \\ &\text{where } \mathfrak{R}_{\mathfrak{f}}\boldsymbol{\xi} \equiv \sum_{k=1}^{d} \Omega_{\mathfrak{f}}(\boldsymbol{\xi}, \mathbf{e}_k)\mathbf{e}_k \end{aligned} \tag{8.58}$$

Finally, as an immediate consequence of (8.50), we have that

$$\mathfrak{R}_{R_O \mathfrak{f}} = O^\top \mathfrak{R}_{\mathfrak{f}} O, \quad \mathfrak{f} \in \mathcal{O}(M) \text{ and } O \in O(\mathbb{R}^d). \tag{8.59}$$

8.4 Derivatives of the Distance Function

In this section we will be examining the *Riemannian distance*

$$\operatorname{dist}^M(x,y) \equiv \inf\left\{ \int_0^1 \|\dot q(t)\|\,dt : q(0) = x \;\&\; q(1) = y \right\} \tag{8.60}$$

between points $x, y \in M$. Just as it was in § 5.3, it is clear that dist^M is a metric for M. In addition, The Hopf–Rinow Theorem, alluded to in § 5.3, says that *completeness of* dist^M *as a metric is equivalent to completeness of* M *as a Riemannian manifold.* Moreover, when M is complete, for each pair of points x and y in the same connected component of M, there is a smooth curve $q : [0,1] \longrightarrow M$ at which the minimum on the right hand side of (8.60) is achieved. Because (cf. § 5.3), after re-parameterization, every such minimizer is a geodesic, we will always take a minimizer for (8.60) to be a geodesic and will call it a *minimal geodesic*. In particular, since $\frac{D\dot q}{dt} \equiv 0$, and therefore

$$\|\dot q(t)\| = \operatorname{dist}^M(x,y), \quad t \in [0,1],$$

for a minimal geodesic, it is an easy matter to see that

$$\operatorname{dist}^M(x,y)^2 = \inf\left\{ \int_0^1 \|\dot q(t)\|^2\,dt : q(0) = x \;\&\; q(1) = y \right\}. \tag{8.61}$$

Our goal is to show how to reproduce in the present setting the reasoning, given in § 5.3.1, which led in § 5.3.2 to Yau's non-explosion criterion (cf. Theorem 5.40).

8.4.1. Yau's Non-Explosion Criterion. Let $x_0 \in M$ be a reference point, define $U(\mathfrak{f}) = \operatorname{dist}^M(\pi\mathfrak{f}, x_0)^2$ for $\mathfrak{f} \in \mathcal{O}(M)$, and, given a smooth, non-decreasing function $F : [0,\infty) \longrightarrow [0,\infty)$, set $g = F \circ U$. If $\mathfrak{e} = (\mathbf{e}_1, \dots, \mathbf{e}_d)$ is an orthonormal basis in $\mathbb{R}^d$, it should now be clear that very few changes are required to make the argument given in § 5.3.1 yield (cf. (5.32))

$$\max_{1 \le k \le d} \left| \int_{\mathcal{O}(M)} \mathfrak{E}(\mathbf{e}_k)^{*\lambda_{\mathcal{O}(M)}} \varphi\, g\, d\lambda_{\mathcal{O}(M)} \right| \le 2 \int_{\mathcal{O}(M)} |\varphi| \sqrt{U}\, F' \circ U\, d\lambda_{\mathcal{O}(M)}$$

for all $\varphi \in C_c^\infty(\mathcal{O}(M);\mathbb{R})$, and (cf. (5.39))

$$\int_{\mathcal{O}(M)} \left(\tfrac{1}{2}\Delta_B \varphi\right) g\, d\lambda_{\mathcal{O}(M)}$$
$$\le \int_{\mathcal{O}(M)} \varphi \left[2U F'' \circ U + d\left(1 + \sqrt{\frac{\kappa U}{d}}\right) F' \circ U \right] d\lambda_{\mathcal{O}(M)}$$

for all $\varphi \in C_c^\infty\big(\mathcal{O}(M);[0,\infty)\big)$ when (cf. (5.35))

$$\kappa(x) \equiv \sup\Big\{-\langle \mathrm{Ric}_y X_y, X_y\rangle : \mathrm{dist}^M(y,x_0) \le \mathrm{dist}^M(x,x_0) \\ \text{and } X_y \in T_yM \text{ with } \|X_y\| \le 1\Big\}.$$

In fact, all that one has to do is systematically replace $D_k^{c,M}$ there with $\mathfrak{E}(\mathbf{e}_k)$ here. To be more precise, in proving the first of these: recall (cf. (8.32)) $\mathfrak{E}(\mathbf{e}_k)^{*\lambda_{\mathcal{O}(M)}} = -\mathfrak{E}(\mathbf{e}_k)$, take $s \rightsquigarrow \mathfrak{p}_k(\,\cdot\,,\mathfrak{f})$ to be the integral curve of $\mathfrak{E}(\mathbf{e}_k)$ with $\mathfrak{p}_k(0,\mathfrak{f}) = \mathfrak{f}$, and repeat the calculation used to prove (5.32), only with $\mathcal{O}(M)$ replacing M and $\mathfrak{p}_k(\,\cdot\,,\mathfrak{f})$ replacing $p_k(\,\cdot\,,x)$. As for the second estimate, one begins by using (8.32) to see, just as in § 5.3.1, that everything comes down to showing that, as $s \searrow 0$, (cf. Lemma 5.34)

$$\sum_{k=1}^d \frac{g\big(\mathfrak{p}_k(s,\mathfrak{f})\big) + g\big(\mathfrak{p}_k(-s,\mathfrak{f})\big) - 2g(\mathfrak{f})}{s^2} \\ \le 4\big(UF''\circ U\big)(\mathfrak{f}) + 2d\big(\beta(x)\coth\beta(x)\big)F'\circ U(\mathfrak{f}) + \mathcal{O}(s)$$

uniformly on compacts. For this purpose, let $x \in M$ and $\mathfrak{f} \in \pi^{-1}(x)$ be given, take $t \in [0,1] \longmapsto q(t,x) \in M$ to be a minimal geodesic from x to x_0, and denote by $\mathfrak{q}(\,\cdot\,,\mathfrak{f})$ the horizontal lift of $q(\,\cdot\,,x)$ with $\mathfrak{q}(0,\mathfrak{f}) = \mathfrak{f}$. Next, for each $1 \le k \le d$, define $s \rightsquigarrow \mathfrak{Q}_k(s,t,\mathfrak{f})$ so that $\mathfrak{Q}_k(\,\cdot\,,t,\mathfrak{f})$ is the integral curve of (cf. the proof of Lemma 5.34) $\alpha(t,x)\mathfrak{E}(\mathbf{e}_k))$ with $\mathfrak{Q}_k(0,t,\mathfrak{f}) = \mathfrak{q}(t,\mathfrak{f})$. Then, just as in the proof of Lemma 5.34, the problem reduces to checking that (5.36) and (5.37) hold when $|\dot{Q}_k(s,t,x)|_{\mathbb{R}^N}$ there is replaced throughout by $\big|\varphi\big(\dot{\mathfrak{Q}}_k(s,t,\mathfrak{f})\big)\big|_{\mathbb{R}^d}$ here. But, by an application of the first structural equation,

$$\frac{d}{ds}\phi\big(\dot{\mathfrak{Q}}_k(s,t,\mathfrak{f})\big) = \dot{\alpha}(t,x)\mathbf{e}_k + \alpha(t,x)\omega\big(\dot{\mathfrak{Q}}_k(s,t,\mathfrak{f})\big)\mathbf{e}_k.$$

In particular, since $\mathfrak{Q}_k(0,\,\cdot\,,\mathfrak{f}) = \mathfrak{q}(\,\cdot\,,\mathfrak{f})$ is horizontal, this leads to the desired analog of (5.36). To complete the program, we have to differentiate the preceding once again with respect to s at $s = 0$, in which case we find, after an application of the second structural equation, that

$$\frac{d^2}{ds^2}\phi\big(\dot{\mathfrak{Q}}_k(s,t,\mathfrak{f})\big)\Big|_{s=0} = \alpha(t,x)\Omega_{\mathfrak{q}(t,\mathfrak{f})}\Big(\mathbf{e}_k, \phi\big(\dot{\mathfrak{q}}(t,\mathfrak{f})\big)\Big)\mathbf{e}_k.$$

But, because $q(\,\cdot\,,x)$ is a geodesic, $\phi\big(\dot{\mathfrak{q}}(t,\mathfrak{f})\big) = \boldsymbol{\theta}(\mathfrak{f}) \equiv \mathfrak{f}^{-1}\dot{q}(0,x)$ for all $t \in [0,1]$. Hence, the rest of the computation is the same as it was in the derivation of the original (5.37).

Using the proceeding, just as we did in §§ 5.3.2 and 5.3.4, we have the following variants of the estimates obtained there.

8.62 THEOREM. *Assume that M is a connected, complete, separable d-dimensional Riemannian manifold, and let $x_0 \in M$ be a fixed reference point. If there exists an $\alpha \in \mathbb{N}$ such that*

$$\langle \mathrm{Ric}_x X_x, X_x \rangle \geq -\frac{\alpha^2}{d}\left(1 + \mathrm{dist}^M(x, x_0)^2\right)\|X_x\|^2$$
$$\text{for all } x \in M \text{ and } X_x \in T_x M,$$

then the martingale problem for $\frac{1}{2}\Delta_B$ on $\mathcal{O}(M)$ is well-posed, and so the martingale problem for $\frac{1}{2}\Delta_M$ on M is also well-posed. In fact, for every $\epsilon \in (0,1)$, $(T, x) \in (0,\infty) \times M$, and $R \in (0,\infty)$:

$$\begin{aligned}(8.63)\qquad &\mathbb{P}_x^M\left(\sup_{t\in[0,T]} \mathrm{dist}^M(p(t), x_0) \geq R\right)\\ &\leq \frac{(1+d+\alpha)!}{(1-\epsilon)^{1+d+\alpha}} \exp\left(\frac{\epsilon\,\mathrm{dist}^M(x,x_0)^2}{2(1-\epsilon)T}\right) e^{-\frac{R^2}{2Te^{\alpha T}}},\end{aligned}$$

and

$$\begin{aligned}(8.64)\qquad &\mathbb{E}^{\mathbb{P}_x^M}\left[\exp\left(\frac{\epsilon\,\mathrm{dist}^M(p(T),x_0)^2}{2Te^{\alpha T}}\right)\right]\\ &\leq \frac{(1+\alpha+d)!}{(1-\epsilon)^{1+d+\alpha}} \exp\left(\frac{\epsilon\,\mathrm{dist}^M(x,x_0)^2}{2(1-\epsilon)T}\right).\end{aligned}$$

Furthermore, if there exists a $\gamma \in \mathbb{R}$ such that

$$\langle \mathrm{Ric}_x X_x, X_x \rangle \geq 2\gamma\|X_x\|^2 \quad \text{for all } x \in M \text{ and } X_x \in T_x M,$$

then

$$\begin{aligned}(8.65)\qquad &\mathbb{P}_x^M\left(\sup_{t\in[0,T]} \mathrm{dist}^M(x, p(t)) \geq R\right)\\ &\leq 2(1-\epsilon)^{-\frac{1}{2}} \exp\left[\epsilon\left(-\frac{R^2}{2T} + \frac{2\dim(M) + \gamma^- \dim(M)^2 T}{1-\epsilon}\right)\right],\end{aligned}$$

for all $R > 0$, and,

$$\begin{aligned}(8.66)\qquad &\mathbb{E}^{\mathbb{P}_x^M}\left[\exp\left(\epsilon\frac{\mathrm{dist}^M(x,p(T))^2}{2T}\right)\right]\\ &\leq 2(1-\epsilon)^{-\frac{1}{2}} \exp\left[\epsilon\frac{2\dim(M) + \gamma^- \dim(M)^2 T}{1-\epsilon}\right].\end{aligned}$$

8.4.2. An Example of Explosion. There would be very little point to our discussing explosion criteria for Brownian motion on complete Riemannian manifolds if we did not know that there are examples for which explosion occurs. For this reason, we will give such examples here, and our examples will show that, in spite of the example in § 5.3.3, the hypothesis in Theorem 8.62 cannot, in general, be substantially weakened.[10]

The examples we have in mind are when $M = \mathbb{R}^2$ and, in the standard, Euclidean coordinate system (x, y), the Riemann metric has the form

$$g_{(x,y)} = \begin{pmatrix} \frac{x^2}{r^2} + \frac{y^2}{r^4}\alpha(r)^2 & -\frac{xy}{r^4}\left(\alpha(r)^2 - r^2\right) \\ -\frac{xy}{r^4}\left(\alpha(r)^2 - r^2\right) & \frac{y^2}{r^2} + \frac{x^2}{r^4}\alpha(r)^2 \end{pmatrix},$$

where $r \equiv \left(x^2 + y^2\right)^{\frac{1}{2}}$ and $\alpha : [0, \infty) \longrightarrow [0, \infty)$ is a function which satisfies $\alpha(r) = r$ for sufficiently small $r \geq 0$ and $\alpha(r) > 0$ for all $r > 0$. In terms of polar coordinates $(r, \theta) \in (0, \infty) \times [0, 2\pi)$ on $\mathbb{R}^2 \setminus \{\mathbf{0}\}$,

$$g_{(r,\theta)} = \begin{pmatrix} 1 & 0 \\ 0 & \alpha(r)^2 \end{pmatrix}.$$

In particular, it is easy to show from this that, no matter how α is chosen, $\text{dist}^M\left(\mathbf{0}, (x, y)\right) = r$ and therefore that bounded balls in M are relatively compact as subsets of $\mathbb{R}^2$. Hence, by the Hopf–Rinow Theorem, every choice of α leads to a complete Riemannian metric. Moreover, straight-forward computation shows that

$$-\text{Ric} = \frac{\alpha''}{\alpha} g. \tag{8.67}$$

In order to test for explosion, we want to look for α's for which there exists a bounded solution to the equation $\frac{1}{2}\Delta_M u = 1$. Indeed, if such a u exists, then for any $R > 0$ and any solution $\mathbb{P}$ to the martingale problem for $\frac{1}{2}\Delta_M$ starting at $\mathbf{0}$,

$$\mathbb{E}^{\mathbb{P}}\left[T \wedge \zeta_R\right] = \mathbb{E}^{\mathbb{P}}\Big[u\big(p(T \wedge \zeta_R\big)\Big] - u(\mathbf{0}), \quad T \geq 0,$$

where $\zeta_R(p)$ is the first time that $p \in \mathcal{P}(\mathbb{R}^2)$ exits from the Euclidean ball of radius R around $\mathbf{0}$. Hence, because u is bounded, we find that $\sup_{R \geq 0} \mathbb{E}^{\mathbb{P}}\left[\zeta_R(p)\right] < \infty$, which certainly means that explosion occurs with probability 1.

---

[10] As I mentioned in the footnote at the beginning of § 4.2.4, in conjunction with the Nash Embedding Theorem (alluded at the beginning of Chapter 7), any reasonable abstract example can be realized as an embedded, closed submanifolds in some $\mathbb{R}^N$. In fact, since our examples below are for $d = 2$, Nash's Theorem says that we can realize our examples as embedded submanifolds of $\mathbb{R}^{51}$. Obviously, for such a simple, explicit examples, it seems clear that one ought to be able to do much better. However, at the moment, I have no idea what is the minimum N for which there exists a closed, embedded submanifold on which the Brownian motion can explode.

To construct α's which have the preceding property, it is easier to work backwards. That is, we start with u and then produce α so that $\Delta_M u = 2$. In fact, we start with $u(x,y) = \psi(r)$, in which case $\Delta_M u = \psi'' + \frac{\alpha'}{\alpha}\psi'$ and therefore we have to take α so that

$$\frac{\alpha'}{\alpha} = \frac{2-\psi''}{\psi'} \text{ on } (0,\infty) \quad \text{and} \quad \alpha(0) = 0.$$

To assure the right behavior of α near 0, we take $\psi(r) = \frac{r^2}{2}$ on $[0,1]$. Next, we extend ψ to the whole of $[0,\infty)$ so that $\psi' > 0$ on $(0,\infty)$, and set $\beta(r) = \frac{1}{\psi'(r)}$ for $r \in [1,\infty)$. In order to make sure that ψ remains bounded, we need to insist that β^{-1} is integrable at infinity. Next, observe that $\frac{\alpha'}{\alpha} = 2\beta + \beta'$, and therefore

$$\frac{\alpha''}{\alpha} = \left(2\beta + \beta'\right)^2 + \left(2\beta+\beta'\right)' = 4\beta^2 + 4\beta\beta' + (\beta')^2 + 2\beta' + \beta''.$$

Hence, for any non-decreasing β with the properties that

$$\int_1^\infty \frac{1}{\beta(r)}\,dr < \infty, \quad \sup\frac{\beta'}{\beta} < \infty, \quad \text{and } \sup\frac{\beta''}{\beta^2} < \infty,$$

we can produce an example of a metric for which the explosion time has finite expected value and yet minus the Ricci curvature grows no faster at infinity than

$$\beta\Big(\operatorname{dist}^M\big(\mathbf{0},(x,y)\big)\Big)^2.$$

For example, we can take $\beta(r) = r\log r$ near infinity, in which case $-\mathrm{Ric}$ grows at infinity like $r^2(\log r)^2$. In this sense, the result in Theorem 8.62 is quite sharp.

8.5 Bochner on $\mathcal{O}(M)$

One of the major advantages gained by working on the bundle of orthonormal frames is that, because one is always specifying a frame, "vectors" and associated quantities on M bear a great deal more resemblance to their realization in the standard Euclidean context. For example, given $F \in C^\infty(\mathcal{O}(M);\mathbb{R})$, define $\mathrm{Grad}^M F : \mathcal{O}(M) \longrightarrow \mathbb{R}^d$ so that

$$(8.68) \qquad \left(\mathrm{Grad}^M_{\mathfrak{f}} F, \boldsymbol{\xi}\right)_{\mathbb{R}^d} = \mathfrak{E}(\boldsymbol{\xi})_{\mathfrak{f}} F, \quad \mathfrak{f} \in \mathcal{O}(M) \text{ and } \boldsymbol{\xi} \in \mathbb{R}^d.$$

Equivalently, given an orthonormal basis $\mathfrak{e} = (\mathbf{e}_1, \dots, \mathbf{e}_d)$ in $\mathbb{R}^d$,

$$\mathrm{Grad}^M F \equiv \sum_{k=1}^d \big(\mathfrak{E}(\mathbf{e}_k)F\big)\mathbf{e}_k.$$

Clearly, for $f \in C_c^\infty(M;\mathbb{R})$,

$$\text{Grad}_{\mathfrak{f}}^M(f\circ\pi) = \mathfrak{f}^{-1}\big(\text{grad}^M f\big)_{\pi\mathfrak{f}}, \tag{8.69}$$

and in this way the gradient can be thought of and treated like a an $\mathbb{R}^d$-valued quantity. Similarly, if we define $\text{Hess}_{\mathfrak{f}}^M F$ so that

$$\big(\boldsymbol{\eta}, \text{Hess}_{\mathfrak{f}}^M \boldsymbol{\xi}\big)_{\mathbb{R}^d} = \mathfrak{E}(\boldsymbol{\eta})_{\mathfrak{f}} \circ \mathfrak{E}(\boldsymbol{\xi})F, \quad \text{for } \mathfrak{f} \in \mathcal{O}(M) \text{ and } \boldsymbol{\xi}, \boldsymbol{\eta} \in \mathbb{R}^d, \tag{8.70}$$

then (cf. (8.30)),

$$\begin{gathered}\langle Y_{\pi\mathfrak{f}}, H_{\pi\mathfrak{f}}^M f X_{\pi\mathfrak{f}}\rangle = \big(\mathfrak{f}^{-1}Y_{\pi\mathfrak{f}}, \text{Hess}_{\mathfrak{f}}^M(f\circ\pi)\mathfrak{f}^{-1}X_{\pi\mathfrak{f}}\big)_{\mathbb{R}^d}\\ \text{for } f \in C^\infty(M;\mathbb{R}) \text{ and } X_{\pi\mathfrak{f}}, Y_{\pi\mathfrak{f}} \in T_{\pi\mathfrak{f}}M.\end{gathered}$$

In particular, the Hessian $\text{Hess}^M(f\circ\pi)$ is a symmetric, $d\times d$-matrix valued function on $\mathcal{O}(M)$.

8.5.1. Bochner's Identity. A beautiful example of the way in which the preceding notation simplifies calculations is provided by the following interpretation and derivation of Bochner's identity (6.19) in the context of $\mathcal{O}(M)$. Namely, Bochner's identity becomes the statement that (cf. (8.29) and (8.58))

$$\text{Grad}^M \Delta_B(f\circ\pi) = \Delta_B \circ \text{Grad}^M(f\circ\pi) - \mathfrak{R}\,\text{Grad}^M(f\circ\pi) \tag{8.71}$$

for $f \in C_c^\infty(M;\mathbb{R})$. Equivalently, and more succinctly,[11]

$$\pi_*\Big(\big[\text{Grad}^M, \Delta_B\big]\Big) = -\mathfrak{R}\,\text{Grad}^M.$$

To check (8.71), note that, for any $\boldsymbol{\xi}, \boldsymbol{\eta} \in \mathbb{R}^d$,

$$\begin{aligned}\mathfrak{E}(\boldsymbol{\xi})\circ\mathfrak{E}(\boldsymbol{\eta})^2(f\circ\pi) &= \mathfrak{E}(\boldsymbol{\eta})\circ\mathfrak{E}(\boldsymbol{\xi})\circ\mathfrak{E}(\boldsymbol{\eta})(f\circ\pi) - \big[\mathfrak{E}(\boldsymbol{\xi}),\mathfrak{E}(\boldsymbol{\eta})\big]\circ\mathfrak{E}(\boldsymbol{\eta})(f\circ\pi)\\ &= \mathfrak{E}(\boldsymbol{\eta})^2\circ\mathfrak{E}(\boldsymbol{\xi})(f\circ\pi) - \mathfrak{E}\big(\Omega(\boldsymbol{\xi},\boldsymbol{\eta})\boldsymbol{\eta}\big)(f\circ\pi),\end{aligned}$$

where, in the passage to the last line, we have used (8.44) twice, once to see that $[\mathfrak{E}(\boldsymbol{\xi}),\mathfrak{E}(\boldsymbol{\eta})](f\circ\pi) = 0$ and again, together with (8.15), to get the second term. To complete the proof from here, replace $\boldsymbol{\xi}$ by $\mathbf{e}_k$, $\boldsymbol{\eta}$ by $\mathbf{e}_\ell$, take both sides of the resulting equality as the coefficient of $\mathbf{e}_k$, and sum the resulting equalities over $1 \le k,\ell \le d$.

---

[11] The reader should not only check that (8.71) is, in fact, the same statement as (6.19) but also that the following expression in terms of a commutation relation reveals that an unexpected cancellation is taking place. Namely, because Grad^M is a first order operation while Δ_B is a second order operator, general principles predict that their commutator will be second order. Thus, it is significant that this commutator turns out to be first order.

8.5.2. Integrated Version of Bochner's Identity. Having obtained (8.71), we will spend the rest of this chapter more or less repeating the reasoning developed in §§ 6.2 and 6.3. Thus, from now on we will be assuming that there exists a $\gamma \in \mathbb{R}$ such that, in the sense of non-negative definite matrices,

$$\mathfrak{R}_{\mathfrak{f}} \geq 2\gamma I \quad \text{for all } \mathfrak{f} \in \mathcal{O}(M). \tag{8.72}$$

Clearly, by Theorem 8.62, (8.72) implies that there is no explosion.

What we are going to do now is repeat the reasoning in § 6.3.2 to get the analog of Theorem 6.16. For this purpose, we will need to have introduced the analog of $\hat{J}_{\mathrm{p}}$ in (6.14); namely, the map $\mathfrak{p} \in \mathcal{P}(\mathcal{O}(M)) \longmapsto \mathfrak{J}_{\mathfrak{p}} \in C^1([0,\infty); \mathrm{Hom}(\mathbb{R}^d, \mathbb{R}^d))$ determined by

$$\dot{\mathfrak{J}}_{\mathfrak{p}}(t) + \tfrac{1}{2}\mathfrak{R}_{\mathfrak{p}(t)}\mathfrak{J}_{\mathfrak{p}}(t) = 0 \quad \text{with } \mathfrak{J}_{\mathfrak{p}}(0) = I. \tag{8.73}$$

Obviously, just as in the derivation of (6.20), (8.72) implies that

$$\big\|\mathfrak{J}_{\mathfrak{p}}(t)\big\| \leq e^{-\gamma t}, \quad (t,\mathfrak{p}) \in [0,\infty) \times \mathcal{P}(\mathcal{O}(M)). \tag{8.74}$$

8.75 LEMMA. *Let $f \in C^\infty(M;\mathbb{R})$ with $\|\mathrm{grad}^M f\|$ bounded be given. Then, $(t,x) \in [0,\infty) \times M \longmapsto [\mathbf{P}_t^M f](x) \in \mathbb{R}$ is smooth and*

$$\big\|[\mathrm{grad}^M \mathbf{P}_t^M f](x)\big\| \leq e^{-\gamma t}\big\|\, \|\mathrm{grad}^M f\|\, \big\|_{C_b(M;\mathbb{R})}, \quad (t,x) \in [0,\infty) \times M.$$

In fact, for each $(T,\mathfrak{f}) \in [0,\infty) \times \mathcal{O}(M)$,

$$\mathrm{Grad}_{\mathfrak{f}}^M \mathbb{P}_T^{\mathcal{O}(M)}(f \circ \pi) = \mathbb{E}^{\mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)}}\Big[\mathfrak{J}_{\mathfrak{p}}(T)^\top \big(\mathrm{Grad}^M(f\circ\pi)\big)_{\mathfrak{p}(T)}\Big]; \tag{8.76}$$

and when $\boldsymbol{\theta} \in C^2([0,T];\mathbb{R}^d)$,

$$\mathfrak{f}\boldsymbol{\theta}(T)\,\mathbf{P}_T^M f - \mathfrak{f}\boldsymbol{\theta}(0)\,\mathbf{P}_T^M f = \mathbb{E}^{\mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)}}\left[\int_0^T \mathfrak{E}\big(\mathfrak{J}_{\mathfrak{p}}(t)\dot{\boldsymbol{\theta}}(t)\big)_{\mathfrak{p}(t)}(f\circ\pi)\,dt\right].$$

PROOF: The smoothness of $(t,x)\rightsquigarrow[\mathbf{P}_t^M f](x)$ is a consequence of Theorem 6.25, and the boundedness of $(t,x)\rightsquigarrow\mathrm{grad}_x^M\mathbf{P}_t f$ comes from (6.30). Given this information, the argument is basically a repeat of the ones given to prove Theorem 6.16 and Lemma 6.37. That is, we set $\mathbf{U}(t,\mathfrak{f}) = \mathrm{Grad}_{\mathfrak{f}}^M \mathbf{P}_{T-t}^{\mathcal{O}(M)}(f\circ\pi)$ for $(t,\mathfrak{f}) \in [0,T] \times \mathcal{O}(M)$. By (8.71),

$$\frac{\partial \mathbf{U}}{\partial t} + \tfrac{1}{2}\Delta_B \mathbf{U} - \tfrac{1}{2}\mathfrak{R}\mathbf{U} = \mathbf{0} \quad \text{on } [0,T] \times \mathcal{O}(M).$$

Hence, by a cut-off procedure and Doob's Stopping Time Theorem, if U is a relatively compact open set in M and

$$\zeta^U(\mathfrak{p}) = \inf\{t \geq 0 : \pi \circ \mathfrak{p}(t) \notin U\},$$

then

$$\mathbf{U}\big(t\wedge T\wedge\zeta^U(\mathfrak{p}),\mathfrak{p}(t\wedge T\wedge\zeta^U)\big)+\int_0^{t\wedge T\wedge\zeta^U(\mathfrak{p})}\tfrac{1}{2}\mathfrak{R}_{\mathfrak{p}(\tau)}\mathbf{U}\big(\tau,\mathfrak{p}(\tau)\big)\,d\tau$$

is an $\mathbb{R}^d$-valued, $\mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)}$-martingale. In particular, by a trivial $\mathbb{R}^d$-valued generalization of Lemma 2.41, this means that

$$\Big(\mathfrak{I}_{\mathfrak{p}}\big(t\wedge T\wedge\zeta^U(\mathfrak{p})\big)\boldsymbol{\theta}\big(t\wedge T\wedge\zeta^U(\mathfrak{p})\big),\mathbf{U}\big(t\wedge T\wedge\zeta^U(\mathfrak{p}),\mathfrak{p}(t\wedge T\wedge\zeta^U)\big)\Big)_{\mathbb{R}^d}$$
$$-\int_0^{t\wedge T\wedge\zeta^U(\mathfrak{p})}\Big(\mathfrak{I}_{\mathfrak{p}}(\tau)\dot{\boldsymbol{\theta}}(\tau),\mathbf{U}\big(\tau,\mathfrak{p}(\tau)\big)\Big)_{\mathbb{R}^d}\,d\tau$$

is an $\mathbb{R}$-valued martingale. Hence, because $(t,x)\rightsquigarrow\|\mathrm{grad}_x^M\mathbf{P}_t^M f\|$ is bounded on $[0,T]\times M$, it is clear that our martingales need not be stopped at ζ^U. That is, we now know that

$$\Big(\mathfrak{I}_{\mathfrak{p}}(t\wedge T)\boldsymbol{\theta}(t\wedge T)),\mathbf{U}\big(t\wedge T,\mathfrak{p}(t\wedge T)\big)\Big)_{\mathbb{R}^d}-\int_0^{t\wedge T}\Big(\mathfrak{I}_{\mathfrak{p}}(\tau)\dot{\boldsymbol{\theta}}(\tau),\mathbf{U}\big(\tau,\mathfrak{p}(\tau)\big)\Big)_{\mathbb{R}^d}\,d\tau$$

is a $\mathbb{P}_{\mathfrak{f}}^{\mathcal{O}(M)}$-martingale for all $T\in(0,\infty)$ and $\boldsymbol{\theta}\in C^2\big([0,\infty);\mathbb{R}^d\big)$. Starting here, the rest is trivial. Namely, the final assertion requires no further comment, whereas (8.76) follows by successively taking $\boldsymbol{\theta}\equiv\mathrm{e}_k$ on $[0,T]$ for $1\le k\le d$ and then summing over k. $\square$

8.5.3. Bismut's Formula on $\mathcal{O}(M)$. With the formulae obtained in Lemma 8.75, we can now prove the analog of (6.43) in the setting of $\mathcal{O}(M)$. The first step is to develop the appropriate form of the Bismut factor (cf. (6.39)) here. Thus, let $\boldsymbol{\theta}\in C^\infty\big([0,\infty);\mathbb{R}^d\big)$ be given. Then, for $(t,\mathfrak{f},\mathbf{w})\in[0,\infty)\times\mathcal{O}(M)\times\mathfrak{W}(\mathbb{R}^d)$, we set

$$B_{\boldsymbol{\theta}}(t,\mathfrak{f},\mathbf{w})=\big[B(\boldsymbol{\theta})\big]\big(t,\mathbf{w},\mathfrak{p}(\,\cdot\,,\mathfrak{f},\mathbf{w})\big)$$
$$\text{where }\big[B(\boldsymbol{\theta})\big](t,\mathbf{w},\mathfrak{p})\equiv\int_0^t\Big(\mathfrak{I}_{\mathfrak{p}}(\tau)\dot{\boldsymbol{\theta}}(\tau),d\mathbf{w}(\tau)\Big)_{\mathbb{R}^d},\tag{8.77}$$

and *the integral is taken in the sense of Riemann–Stieltjes.* That is, because, for any $\mathfrak{p}\in\mathcal{P}\big(\mathcal{O}(M)\big)$, $\tau\longmapsto\mathfrak{I}_{\mathfrak{p}}(\tau)\boldsymbol{\theta}(\tau)$ is continuously differentiable and therefore

$$\int_0^t\left(\frac{d\mathfrak{I}_{\mathfrak{p}}(\tau)\dot{\boldsymbol{\theta}}(\tau)}{d\tau},\mathbf{w}(\tau)\right)_{\mathbb{R}^d}d\tau$$

is well-defined as a Riemann integral, the existence of the right hand side of (8.77) as a Riemann–Stieltjes integral follows from integration by parts (cf. Theorem 1.2.7 in [36]). In particular, since (cf. (3.17)) $\mathfrak{p}(\,\cdot\,,*,\mathbf{w}_n)\longrightarrow$

$\mathfrak{p}(\cdot\,,*,\mathbf{w})$ uniformly on compact subsets of $[0,\infty)\times\mathcal{O}(M)$ for every $\mathbf{w}\in\mathfrak{W}(\mathbb{R}^d)$, we also know that

$$\begin{bmatrix}\mathfrak{p}(t,\mathfrak{f},\mathbf{w}_n)\\ \mathfrak{J}_{\mathfrak{p}(\cdot\,,\mathfrak{f},\mathbf{w}_n)}(t)\\ B_{\boldsymbol{\theta}}(t,\mathfrak{f},\mathbf{w}_n)\end{bmatrix}\longrightarrow\begin{bmatrix}\mathfrak{p}(t,\mathfrak{f},\mathbf{w})\\ \mathfrak{J}_{\mathfrak{p}(\cdot\,,\mathfrak{f},\mathbf{w})}(t)\\ B_{\boldsymbol{\theta}}(t,\mathfrak{f},\mathbf{w})\end{bmatrix}\tag{8.78}$$

uniformly on compact subsets of $[0,\infty)\times\mathcal{O}(M)$ for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$. In fact, if, for $A\in\mathrm{Hom}(\mathbb{R}^d;\mathbb{R}^d)$, $\rho(A)$ is the vector field on $\mathrm{Hom}(\mathbb{R}^d;\mathbb{R}^d)$ determined at $J\in\mathrm{Hom}(\mathbb{R}^d;\mathbb{R}^d)$ by $\frac{d}{ds}e^{sA}J\big|_{s=0}$, and, for a given choice of orthonormal basis $\mathfrak{e}=(\mathbf{e}_1,\dots,\mathbf{e}_d)$ in $\mathbb{R}^d$, $\mathfrak{Y}_k^{\mathfrak{e},\mathcal{O}(M)}(t)$ is defined on $\mathcal{O}(M)\times\mathrm{Hom}(\mathbb{R}^d;\mathbb{R}^d)\times\mathbb{R}$ so that

$$\big(\mathfrak{Y}_k^{\mathfrak{e},\mathcal{O}(M)}(t)\big)_{(\mathfrak{f},J,B)}\equiv\mathfrak{E}(\mathbf{e}_k)_{\mathfrak{f}}+\big(J\dot{\boldsymbol{\theta}}(t),\mathbf{e}_k\big)_{\mathbb{R}^d}\partial_B\quad\text{for } 1\le k\le d,$$

then (just as in § 6.3.2), for each $\mathfrak{f}\in\mathcal{O}(M)$ the $\mu_{\mathbb{R}^d}$-distribution of

$$\mathbf{w}\in\mathfrak{W}(\mathbb{R}^d)\longmapsto\begin{bmatrix}\mathfrak{p}(\,*\,,\mathfrak{f},\mathbf{w})\\ \mathfrak{J}_{\mathfrak{p}(\cdot\,,\mathfrak{f},\mathbf{w})}(\,*\,)\\ B_{\boldsymbol{\theta}}(\,*\,,\mathfrak{f},\mathbf{w})\end{bmatrix}\in C\big([0,\infty);\mathcal{O}(M)\times\mathrm{Hom}(\mathbb{R}^d;\mathbb{R}^d)\times\mathbb{R}\big)$$

is uniquely characterized by the fact that it solves the martingale problem starting at $(\mathfrak{f},I,0)$ for the time-dependent operator

$$t\rightsquigarrow-\frac{1}{2}\rho(\mathfrak{R}_{\mathfrak{f}})_J+\frac{1}{2}\sum_{k=1}^d\big(\mathfrak{Y}_k^{\mathfrak{e},\mathcal{O}(M)}(t)\big)^2_{(\mathfrak{f},J,B)}.$$

Hence, by exactly the argument given to prove Theorem 6.41, we have that

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[e^{\alpha B_{\boldsymbol{\theta}}(T,\mathfrak{f},\mathbf{w})}\Big]\le\exp\left(\frac{\alpha^2}{2}\int_0^T e^{-2\gamma t}|\dot{\boldsymbol{\theta}}(t)|^2_{\mathbb{R}^d}\,dt\right)\tag{8.79}$$

and, for any $f\in C_{\mathrm{b}}(M;\mathbb{R})$,

$$\mathfrak{f}\boldsymbol{\theta}(T)\,\mathbf{P}_T^Mf-\mathfrak{f}\boldsymbol{\theta}(0)\mathbf{P}_T^Mf=\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[B_{\boldsymbol{\theta}}(T,\mathfrak{f},\mathbf{w})f\circ\pi\big(\mathfrak{p}(T,\mathfrak{f},\mathbf{w})\big)\Big],\tag{8.80}$$

the latter being the form that *Bismut's formula* takes in $\mathcal{O}(M)$.

Obviously, once (8.80) is known, there is nothing preventing us from simply repeating the line of reasoning which led to the estimates in Theorem 6.46. which we therefore now know hold in the present context as well.

8.5.4. A Technical Comment. The careful reader may have noticed that there appears to be a peculiar distinction between the Bismut factor in (8.77) and the one in (6.38) and (6.39). Namely, because of the appearance of the non-differentiable factor $t\rightsquigarrow O(t,x,\mathbf{w})$ in the integrand of (6.38), we could not use simple Riemann–Stieltjes integration to define that integral and had to resort an almost sure approximation procedure. This is in stark contrast to the definition of the Bismut factor in (8.77) which admits a path-by-path definition in terms of Riemann–Stieltjes integration. Compounding this confusion, at least for those who know about such technicalities, is the question whether the $B_{\boldsymbol{\theta}}(t,x,\mathbf{w})$ in (6.39) is actually an Itô or a Stratonovich integral. Indeed, the definition which we gave there indicates that it is a Stratonovich integral, whereas the estimate in (6.42) indicates that it is an Itô integral. This problem does not arise for the Bismut factor in (8.77) because Itô's theory coincides with Stratonovich's when applied to integrals which can be interpreted in terms of Riemann–Stieltjes theory.

There are two explanations for what is going on. In the first place, the absence from (8.77) of a factor like $O(t,x,\mathbf{w})$ is, in a sense, illusory. The point is that it is hidden in $\dot{\mathbf{w}}(t)$ by the fact that, in our present construction, the "velocity" $\dot{\mathfrak{p}}(t,\mathfrak{f},\mathbf{w})$ of the Brownian path is always expressed in terms of the frame $\mathfrak{p}(t,\mathfrak{f},\mathbf{w})$ which has evolved along the Brownian path $p(\,\cdot\,,\pi\mathfrak{f},\mathbf{w})$ via parallel transport. In our earlier construction, we were working in a fixed frame, and the appearance of the orthogonal transformation $O(t,x,\mathbf{w})$ is precisely what was required to compensate for this fact. Secondly, turning to the question about Itô versus Stratonovich, the preceding geometric observation is reflected in the stochastic observation that, even though the Bismut factor in (6.39) is not amenable to Riemann–Stieltjes theory, Itô's and Stratonovich's theories give the same answer in that computation. Those readers who are schooled in the theory of stochastic integration will realize that this coincidence is a stochastic manifestation of the identity (of which we have already taken frequent advantage) $\sum_{k=1}^{N}\partial\Pi^M(D_k^{\mathfrak{c},M})(\Pi^M)^{\perp}\mathbf{e}_k \equiv \mathbf{0}$.

CHAPTER 9

Local Analysis of Brownian Motion

The main purpose of this chapter is to provide a brief introduction to some basic techniques which are available for refined analysis of the way Brownian motion on a Riemannian manifold M responds to the local geometry of M. Throughout, we will assume that M is separable, complete, and connected and that the Brownian motion on M never explodes. In fact, because we will be mostly dealing with local issues, little or nothing would be lost if we were to assume that M is compact.

Since everything we will be doing relies on them, we will begin by reviewing normal coordinate systems.

9.1 Normal Coordinates

Given $\mathfrak{f} \in \mathcal{O}(M)$, define $\operatorname{Exp}_{\mathfrak{f}} : \mathbb{R}^d \longrightarrow M$ so that $\operatorname{Exp}_{\mathfrak{f}}(\mathbf{0}) = \mathfrak{f}$ and, for any $\boldsymbol{\xi} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$,

$$\frac{d}{dt}\operatorname{Exp}_{\mathfrak{f}}(t\boldsymbol{\xi}) = \mathfrak{E}(\boldsymbol{\xi})_{\operatorname{Exp}_{\mathfrak{f}}(t\boldsymbol{\xi})}, \quad t \in [0,\infty). \tag{9.1}$$

Then, for $x \in M$ and $\mathfrak{f} \in \pi^{-1}x$, the map $\exp_{\mathfrak{f}} : \mathbb{R}^d \longrightarrow M$ given by

$$\exp_{\mathfrak{f}}(\boldsymbol{\theta}) = \pi \circ \operatorname{Exp}_{\mathfrak{f}}(\boldsymbol{\theta}) \tag{9.2}$$

is called the *exponential map* based at x in the frame $\mathfrak{f}$. Clearly (cf. § 8.1.2), $\exp_{\mathfrak{f}}(\boldsymbol{\theta})$ is the position at time 1 of the (unique) geodesic which starts from x with initial velocity $\mathfrak{f}\boldsymbol{\theta}$. Thus, for each $\mathbf{e} \in \mathbf{S}^{d-1}M$, there is an $R_{\mathfrak{f}}(\mathbf{e}) \in (0,\infty]$ such that

$$\begin{aligned} &\operatorname{dist}^M\big(x, \exp_{\mathfrak{f}}(r\mathbf{e})\big) = r \quad \text{for all } 0 \le r < R_{\mathfrak{f}}(\mathbf{e}) \\ &\operatorname{dist}^M\big(x, \exp_{\mathfrak{f}}(r\mathbf{e})\big) < r \quad \text{for all } R_{\mathfrak{f}}(\mathbf{e}) < r < \infty. \end{aligned} \tag{9.3}$$

If $R_{\mathfrak{f}}(\mathbf{e}) < \infty$, then $\exp_{\mathfrak{f}}\big(R_{\mathfrak{f}}(\mathbf{e})\mathbf{e}\big)$ is called the *cut point* along the geodesic ray $t \rightsquigarrow \exp_{\mathfrak{f}}(t\mathbf{e})$ in the direction $\mathfrak{f}\mathbf{e}$, and the *cut locus* $\operatorname{Cut}(x)$ of x is the set of cut points along geodesic rays emanating from x. Set

$$\mathcal{U}_{\mathfrak{f}} = \big\{r\mathbf{e} : \mathbf{e} \in \mathbf{S}^{d-1} \text{ and } 0 \le r < R_{\mathfrak{f}}(\mathbf{e})\big\}.$$

Then, elementary analysis (cf. the discussion in § 2 of Chapter 13 in [8]) shows that the map $\boldsymbol{\theta} \in \mathcal{U}_{\mathfrak{f}} \longmapsto \exp_{\mathfrak{f}}(\boldsymbol{\theta}) \in M$ is diffeomorphic onto $M \setminus \mathrm{Cut}(x)$. Thus, if

$$N_x = \exp_{\mathfrak{f}}(\mathcal{U}_{\mathfrak{f}}) \text{ and } \Theta_{\mathfrak{f}} = \exp_{\mathfrak{f}}^{-1} \restriction N_x, \tag{9.4}$$

then N_x is an open neighborhood of x and $(N_x, \Theta_{\mathfrak{f}})$ is a coordinate chart. Coordinate systems which arise in this way play such an essential role in all differential geometric analysis that they are called *normal coordinate systems.* One of reasons why they are so important is that, at the point x, they behave like standard Euclidean coordinates on $\mathbb{R}^d$. More precisely,

$$\nabla_{(\partial_{\boldsymbol{\theta}}^{\Theta_{\mathfrak{f}}})_x} \partial_{\boldsymbol{\eta}}^{\Theta_{\mathfrak{f}}} = \mathbf{0}, \quad \text{for all } \mathfrak{f} \in \pi^{-1} \text{ and } \boldsymbol{\theta},\ \boldsymbol{\eta} \in \mathbb{R}^d. \tag{9.5}$$

To see this, first note that, by an obvious polarization argument, it suffices to check the case when $\boldsymbol{\eta} = \boldsymbol{\theta}$. But in this case, $(\partial^{\Theta_{\mathfrak{f}}})_{\exp_{\mathfrak{f}}(t\boldsymbol{\theta})}$ is obtained from $(\partial^{\Theta_{\mathfrak{f}}})_x$ by parallel transport along $\tau \in [0,t] \longmapsto \exp_{\mathfrak{f}}(\tau\boldsymbol{\theta}) \in M$, which, by (5.11), leads immediately to (9.5) when $\boldsymbol{\eta} = \boldsymbol{\theta}$.

9.1.1. Relationship to the Distance Function. In this subsection, we show that the coordinate map $\Theta_{\mathfrak{f}}$ is intimately related to the distance function based at $\pi\mathfrak{f}$.

9.6 LEMMA. Set

$$U(\mathfrak{f}, y) = \mathrm{dist}^M\big(\pi(\mathfrak{f}), y\big)^2 \quad \text{for } (\mathfrak{f}, y) \in \mathcal{O}(M) \times M.$$

If (cf. (9.4)) $y \in N_x$, then for each $\mathfrak{f} \in \pi^{-1}x$,

$$\Theta_{\mathfrak{f}}(y) = -\tfrac{1}{2}\mathrm{Grad}_{\mathfrak{f}}^M U(\,\cdot\,, y) = \tfrac{1}{2}\mathrm{Grad}_{\mathfrak{F}_{\mathfrak{f}}(y)}^M U\big(\pi(\mathfrak{f}), \cdot\,\big), \tag{9.7}$$

where

$$\mathfrak{F}_{\mathfrak{f}}(y) \equiv \mathrm{Exp}_{\mathfrak{f}}\big(\Theta_{\mathfrak{f}}(y)\big). \tag{9.8}$$

PROOF: Let $\mathfrak{f} \in \pi^{-1}x$ and $\boldsymbol{\xi} \in \mathbb{R}^d$ be given, and set $\mathfrak{f}_s = \mathrm{Exp}_{\mathfrak{f}}(s\boldsymbol{\xi})$ for $s \in \mathbb{R}$. Then, because $y \in N_x$, there exists a $\delta > 0$ such that $y \in N_{\pi\mathfrak{f}_s}$ for all $|s| < \delta$. Hence, if we define $\mathfrak{p}_s(t) = \mathrm{Exp}_{\mathfrak{f}_s}\big(t\Theta_{\mathfrak{f}_s}(y)\big)$ for $(s,t) \in (-\delta,\delta) \times [0,1]$, then

$$U(\mathfrak{f}_s, y) = \int_0^1 \big|\phi(\dot{\mathfrak{p}}_s(t))\big|_{\mathbb{R}^d}^2\, dt$$

for each $s \in (-\delta, \delta)$. Thus

$$\frac{d}{ds}U(\mathfrak{f}_s, y) = 2\int_0^1 \Big(\tfrac{d}{ds}\phi\big(\dot{\mathfrak{p}}_s(t)\big), \phi\big(\dot{\mathfrak{p}}_s(t)\big)\Big)_{\mathbb{R}^d}\, dt.$$

But, by Cartan's first structural equation (8.48),

$$\tfrac{d}{ds}\phi(\dot{\mathfrak{p}}_s(t)) = \tfrac{d}{dt}\phi(\mathfrak{p}'_s(t)) - \omega(\mathfrak{p}'_s(t))\phi(\dot{\mathfrak{p}}_s(t))$$

since $\dot{\mathfrak{p}}_s(t)$ is horizontal. Further, because $\omega(\mathfrak{p}'_s(t))$ is skew symmetric, we now see that

$$\frac{d}{ds}U(\mathfrak{f}_s, y) = 2\int_0^1 \Big(\tfrac{d}{dt}\phi(\mathfrak{p}'_s(t)), \phi(\dot{\mathfrak{p}}_s(t))\Big)_{\mathbb{R}^d}\, dt.$$

But $\phi(\dot{\mathfrak{p}}_s(t)) = \Theta_{\mathfrak{f}_s}(y)$ for all $t \in [0,1]$, and therefore the integral in the preceding can be performed and yields

$$\frac{d}{ds}U(\mathfrak{f}_s, y)\Big|_{s=0} = 2\Big(\phi(\mathfrak{p}'_0(t)), \Theta_{\mathfrak{f}}(y)\Big)_{\mathbb{R}^d}\Big|_{t=0}^{1} = -2\big(\boldsymbol{\xi}, \Theta_{\mathfrak{f}}(y)\big)_{\mathbb{R}^d},$$

since $\phi(\mathfrak{p}'_0(t))$ equals $\boldsymbol{\xi}$ at $t = 0$ and $\mathbf{0}$ at $t = 1$. In particular, we have now shown that

$$\mathfrak{E}(\boldsymbol{\xi})_{\mathfrak{f}}U(\,\cdot\,, y) = -2\big(\boldsymbol{\xi}, \Theta_{\mathfrak{f}}(y)\big)_{\mathbb{R}^d}, \quad \boldsymbol{\xi} \in \mathbb{R}^d,$$

which is equivalent to the first equality in (9.7). Given the first equality, the second equality follows immediately from the essentially obvious observation that $\Theta_{\mathfrak{f}}(y) = -\Theta_{\mathfrak{F}_{\mathfrak{f}}(y)}(x)$. Indeed, if $p : [0,1] \longrightarrow M$ is the minimal geodesic running from x to y, then $t \rightsquigarrow p(1-t)$ is the minimal geodesic from y to x. Hence $\Theta_{\mathfrak{g}}(x) = -\mathfrak{g}^{-1}\dot{p}(1)$ for any $\mathfrak{g} \in \pi^{-1}y$. But, $\dot{p}(1) = \mathcal{T}_p\dot{p}(0)$, and therefore $\dot{p}(1) = \mathfrak{F}_{\mathfrak{f}}(y)\Theta_{\mathfrak{f}}(y)$. □

Next, given $\mathfrak{f} \in \mathcal{O}(M)$ and $\mathfrak{g} \in \pi^{-1}(N_{\pi\mathfrak{f}})$, define $\Sigma(\mathfrak{f},\mathfrak{g}) \in \mathrm{Hom}(\mathbb{R}^d;\mathbb{R}^d)$ so that

$$\Sigma(\mathfrak{f},\mathfrak{g}) = \mathfrak{E}(\boldsymbol{\eta})_{\mathfrak{g}}\Theta_{\mathfrak{f}} \circ \pi. \tag{9.9}$$

9.10 Lemma. *Given $x \in M$ and $\mathfrak{f} \in \pi^{-1}x$, (cf. (9.8))*

$$\Sigma(\mathfrak{f},\mathfrak{g}) = \Sigma\big(\mathfrak{f}, \mathfrak{F}_{\mathfrak{f}}(y)\big) \circ \mathfrak{F}_{\mathfrak{f}}(y)^{-1} \circ \mathfrak{g} \quad \text{for any } \mathfrak{g} \in \pi^{-1}N_x.$$

Further, if $y \in N_x \longmapsto g_y^{\Theta_{\mathfrak{f}}} \in \mathrm{Hom}(\mathbb{R}^d;\mathbb{R}^d)$ is the Riemann metric in terms of the coordinate system $(N_x, \Theta_{\mathfrak{f}})$, then

$$g_y^{\Theta_{\mathfrak{f}}} = \big(\Sigma(\mathfrak{f},\mathfrak{g})\Sigma(\mathfrak{f},\mathfrak{g})^\top\big)^{-1} \quad \text{for } y \in N_x \text{ and } \mathfrak{g} \in \pi^{-1}y. \tag{9.11}$$

Proof: To see the first part, simply note that

$$\begin{aligned}\Sigma(\mathfrak{f},\mathfrak{g})\boldsymbol{\eta} = \mathfrak{E}(\boldsymbol{\eta})_{\mathfrak{g}}\Theta_{\mathfrak{f}} \circ \pi &= \mathfrak{E}\big(\mathfrak{F}_{\mathfrak{f}}(y)^{-1} \circ \mathfrak{g}\boldsymbol{\eta}\big)_{\mathfrak{F}_{\mathfrak{f}}(y)}\Theta_{\mathfrak{f}} \circ \pi \\ &= \Sigma\big(\mathfrak{f}, \mathfrak{F}_{\mathfrak{f}}(y)\big)\mathfrak{F}_{\mathfrak{f}}(y)^{-1} \circ \mathfrak{g}\boldsymbol{\eta}.\end{aligned}$$

Turning to the second part, let $\boldsymbol{\eta} \in \mathbb{R}^d$ be given. By definition, $(\partial_{\boldsymbol{\eta}}^{\Theta_{\mathfrak{f}}})_y = (\Theta_{\mathfrak{f}}^{-1})_*(\partial_{\boldsymbol{\eta}})_{\Theta_{\mathfrak{f}}(y)}$. Hence

$$\boldsymbol{\eta} = (\partial_{\boldsymbol{\eta}}^{\Theta_{\mathfrak{f}}})_y\Theta_{\mathfrak{f}} = \Sigma(\mathfrak{f},\mathfrak{g})\mathfrak{g}^{-1}(\partial_{\boldsymbol{\eta}}^{\Theta_{\mathfrak{f}}})_y,$$

and so $\mathfrak{g}^{-1}(\partial_{\eta}^{\Theta_{\mathfrak{f}}})_y = \Sigma(\mathfrak{f},\mathfrak{g})^{-1}\eta$. In particular,

$$\begin{aligned}(\zeta, g_y^{\Theta_{\mathfrak{f}}}\eta)_{\mathbb{R}^d} &= \langle(\partial_{\zeta}^{\Theta_{\mathfrak{f}}})_y, (\partial_{\eta}^{\Theta_{\mathfrak{f}}})_y\rangle \\ = \left(\Sigma(\mathfrak{f},\mathfrak{g})^{-1}\zeta, \Sigma(\mathfrak{f},\mathfrak{g})^{-1}\eta\right)_{\mathbb{R}^d} &= \left(\zeta, \left(\Sigma(\mathfrak{f},\mathfrak{g})\Sigma(\mathfrak{f},\mathfrak{g})^{\top}\right)^{-1}\eta\right)_{\mathbb{R}^d}. \qquad \square\end{aligned}$$

9.2 Brownian Motion in Normal Coordinates

When Itô first attempted to construct Brownian motion on a Riemannian manifold, he did so by a laborious procedure which required him to patch together constructions carried out in local coordinate charts. Thus, it is interesting, if potentially irresponsible, to examine just how close he could have come to getting away with just one normal coordinate chart.

To develop this point, let x and $\mathfrak{f} \in \pi^{-1}x$ be given. If $\mathbf{w}$ is piecewise smooth and (cf. (8.27) and (8.37)) $p(t,\mathfrak{f},\mathbf{w}) \in N_x$, then (cf. the first part of Lemma 9.10)

$$\frac{d}{dt}\Theta_{\mathfrak{f}}\left(p(t,\mathfrak{f},\mathbf{w})\right) = \Sigma\left(\mathfrak{f},\mathfrak{p}(t,\mathfrak{f},\mathbf{w})\right)\dot{\mathbf{w}}(t) = \Sigma\Big(\mathfrak{f},\mathfrak{F}_{\mathfrak{f}}\left(p(t,\mathfrak{f},\mathbf{w})\right)\Big)O(t,\mathfrak{f},\mathbf{w})\dot{\mathbf{w}}(t)$$
$$\text{where } O(t,\mathfrak{f},\mathbf{w}) \equiv \mathfrak{F}_{\mathfrak{f}}\left(p(t,x,\mathbf{w})\right)^{-1} \circ \mathfrak{p}(t,\mathfrak{f},\mathbf{w}).$$

At the same time, since $\mathfrak{p}(t,\mathfrak{f},\mathbf{w}) = R_{O(t,\mathfrak{f},\mathbf{w})}\mathfrak{F}_{\mathfrak{f}}\left(p(t,\mathfrak{f},\mathbf{w})\right)$ and $\mathfrak{p}(\,\cdot\,,\mathfrak{f},\mathbf{w})$ is horizontal,

$$\begin{aligned}&\lambda\left(O(t,\mathfrak{f},\mathbf{w})^{\top}\dot{O}(t,\mathfrak{f},\mathbf{w})\right)_{\mathfrak{p}(t,\mathfrak{f},\mathbf{w})} \\ &\qquad = -\left(R_{O(t,\mathfrak{f},\mathbf{w})}\right)_*\lambda\Big(\omega\Big(\tfrac{d}{dt}\mathfrak{F}_{\mathfrak{f}}\left(p(t,\mathfrak{f},\mathbf{w})\right)\Big)\Big)_{\mathfrak{F}_{\mathfrak{f}}(p(t,\mathfrak{f},\mathbf{w}))},\end{aligned}$$

from which we see, after an application of (8.5), that

$$\dot{O}(t,\mathfrak{f},\mathbf{w}) = -\omega\Big(\tfrac{d}{dt}\mathrm{Exp}_{\mathfrak{f}}\left(\boldsymbol{\theta}(t,\mathfrak{f},\mathbf{w})\right)\Big)O(t,\mathfrak{f},\mathbf{w})$$
$$\text{where } \boldsymbol{\theta}(t,\mathfrak{f},\mathbf{w}) \equiv \Theta_{\mathfrak{f}}\left(p(t,\mathfrak{f},\mathbf{w})\right).$$

By combining these, we now know that, as long as $p(t,\mathfrak{f},\mathbf{w}) \in N_x$,

$$\begin{aligned}\dot{\boldsymbol{\theta}}(t,\mathfrak{f},\mathbf{w}) &= \Sigma\Big(\mathfrak{f},\mathrm{Exp}_{\mathfrak{f}}\left(\boldsymbol{\theta}(t,\mathfrak{f},\mathbf{w})\right)\Big)O(t,\mathfrak{f},\mathbf{w})\dot{\mathbf{w}}(t) \\ \dot{O}(t,\mathfrak{f},\mathbf{w}) &= -\omega\Big(\tfrac{d}{dt}\mathrm{Exp}_{\mathfrak{f}}\left(\boldsymbol{\theta}(t,\mathfrak{f},\mathbf{w})\right)\Big)O(t,\mathfrak{f},\mathbf{w}).\end{aligned}$$

Hence, if $(\mathbf{e}_1,\dots,\mathbf{e}_d)$ is an orthonormal basis in $\mathbb{R}^d$ and we define the vector fields $Z_1,\dots,Z_d$ on $\Theta_{\mathfrak{f}}(N_x)\times O(\mathbb{R}^d)$ so that

$$(9.12)\quad (Z_k)_{(\boldsymbol{\theta},O)} = \sum_{i=1}^{d}\left(\mathbf{e}_i, \Sigma(\mathfrak{f},\mathrm{Exp}_{\mathfrak{f}}(\boldsymbol{\theta}))O\mathbf{e}_k\right)_{\mathbb{R}^d}\left(\left(\partial_i\right)_{\boldsymbol{\theta}} - \rho\Big(\omega\big((\partial_i)_{\boldsymbol{\theta}}\mathrm{Exp}_{\mathfrak{f}}\big)\Big)_O\right),$$

where (cf. § 5.1.2) we use $\rho(a)$ to denote the right-invariant vector field on $O(\mathbb{R}^d)$ determined by $a \in o(\mathbb{R}^d)$, then the preceding system is equivalent to

$$\frac{d}{dt}\begin{pmatrix}\boldsymbol{\theta}(t,\mathfrak{f},\mathbf{w})\\ O(t,\mathfrak{f},\mathbf{w})\end{pmatrix} = \sum_{k=1}^{d}\big(\mathbf{e}_k,\dot{\mathbf{w}}(t)\big)_{\mathbb{R}^d}(Z_k)_{(\boldsymbol{\theta}(t,\mathfrak{f},\mathbf{w}),O(t,\mathfrak{f},\mathbf{w}))} \tag{9.13}$$
$$\text{for } t \notin \mathrm{Cut}(\mathfrak{f},\mathbf{w}) \equiv \big\{\tau \ge 0 : p(\tau,\mathfrak{f},\mathbf{w}) \in \mathrm{Cut}(\pi\mathfrak{f})\big\}.$$

At first sight, (9.13) looks pretty hopeful. Indeed, because $\mathrm{Cut}(\pi\mathfrak{f})$ is closed, so is $\mathrm{Cut}(\mathfrak{f},\mathbf{w})$. More important, because $\mathrm{Cut}(\pi\mathfrak{f})$ has λ_M-measure 0 (cf. Proposition 3.1 in [5]), $\mu_{\mathbb{R}^d}$-almost no $p(\,\cdot\,,\mathfrak{f},\mathbf{w})$ spends positive time in $\mathrm{Cut}(\pi\mathfrak{f})$ and therefore $\mathrm{Cut}(\mathfrak{f},\mathbf{w})$ has measure 0. To see this, let $g_t(x,y)$ denote the heat kernel (i.e., minimal fundamental solution) for $\partial_t u = \frac{1}{2}\Delta_M u$, and apply Fubini's Theorem to justify

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\int_0^\infty \mathbf{1}_{\mathrm{Cut}(\pi\mathfrak{f})}\big(p(\tau,\mathfrak{f},\mathbf{w})\big)\,d\tau\right] = \int_0^\infty\left(\int_{\mathrm{Cut}(\pi\mathfrak{f})} g_\tau(\pi\mathfrak{f},y)\,\lambda_M(dy)\right)d\tau = 0.$$

From our point of view, the immediate importance of these observations is the conclusion that, for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$, $\mathfrak{p}(\,\cdot\,,\mathfrak{f},\mathbf{w})$ is determined by its own restriction to $[0,\infty)\setminus\mathrm{Cut}(\mathfrak{f},\mathbf{w})$; and, as long as $t \notin \mathrm{Cut}(\mathfrak{f},\mathbf{w})$,

$$\mathfrak{p}(t,\mathfrak{f},\mathbf{w}) = R_{O(t,\mathfrak{f},\mathbf{w})}\mathrm{Exp}_{\mathfrak{f}}\big(\boldsymbol{\theta}(t,\mathfrak{f},\mathbf{w})\big).$$

Finally, on $[0,\infty)\setminus\mathrm{Cut}(\mathfrak{f},\mathbf{w})$, one might hope to be able to construct the pair $\big(\boldsymbol{\theta}(\,\cdot\,,\mathfrak{f},\mathbf{w}),O(\,\cdot\,,\mathfrak{f},\mathbf{w})\big)$ by an approximation procedure in which $\mathbf{w}$ in (9.13) is replaced by $\mathbf{w}_n$ and n is allowed to tend to infinity. Indeed, there is no doubt that this scheme will work during the initial interval $\big[0,\zeta(\mathfrak{f},\mathbf{w})\big)$, where $\zeta(\mathfrak{f},\mathbf{w})$ denotes the first time that $p(\,\cdot\,,\mathfrak{f},\mathbf{w})$ visits $\mathrm{Cut}(\pi\mathfrak{f})$. The problem comes at time $\zeta(\mathfrak{f},\mathbf{w})$. Namely, because (9.13) breaks down at $\mathrm{Cut}(\pi\mathfrak{f})$, we can not use it to continue until the path gets away from $\mathrm{Cut}(\pi\mathfrak{f})$, and (9.13) provides no information about where the path will be when it does get away. Of course, at time $\zeta(\mathfrak{f},\mathbf{w})$, we can switch from the normal coordinate system based at $\mathfrak{f}$ to the one at

$$\mathfrak{p}\big(\zeta(\mathfrak{f},\mathbf{w}),\mathfrak{f},\mathbf{w}\big) = \lim_{t\nearrow\zeta(\mathfrak{f},\mathbf{w})} R_{O(t,\mathfrak{f},\mathbf{w})}\mathrm{Exp}_{\mathfrak{f}}\big(\boldsymbol{\theta}(t,\mathfrak{f},\mathbf{w})\big)$$

and start all over again. But, in the final analysis, this is more or less what Itô was forced to do also.

9.3 Asymptotic Expansion of Metric in Normal Coordinates

As we mentioned earlier, a distinguishing feature of normal coordinate systems is the property expressed by (9.5). A formulation of this property with which Riemann himself was familiar[1] is the statement that first order derivatives of the metric $g^{\Theta_{\mathfrak{f}}}$ vanish at the point $x = \pi\mathfrak{f}$. Indeed, (9.5) implies that

$$(\partial_{\mathbf{e}_k}^{\Theta_{\mathfrak{f}}})_x(\partial_{\mathbf{e}_i}^{\Theta_{\mathfrak{f}}}, \partial_{\mathbf{e}_j}^{\Theta_{\mathfrak{f}}}) = 0 \quad, 1 \le i, j, k \le d.$$

Hence, $g^{\Theta_{\mathfrak{f}}}y - I_{\mathbb{R}^d} = \mathcal{O}(r^2)$ where $r = \mathrm{dist}^M(y,x)$. Following Riemann, we devote this section to showing that this $\mathcal{O}(r^2)$ term contains a good deal of geometric information.

9.3.1. Relationship to Jacobi Fields. The basic tool which will enable to carry out this program is contained in the following lemma.

9.14 LEMMA. *Let (cf. (9.4)) $y \in N_x$ and $\mathfrak{f} \in \pi^{-1}x$ be given, and set $\mathfrak{p}(t) = \mathrm{Exp}_{\mathfrak{f}}\big(t\Theta_{\mathfrak{f}}(y)\big)$ for $t \in [0,1]$. Then for each $\boldsymbol{\eta} \in \mathbb{R}^d$ there is a unique smooth $\mathbf{J}_{\boldsymbol{\eta}} : [0,1] \longrightarrow \mathbb{R}^d$ satisfying*

$$\text{(9.15)} \quad \ddot{\mathbf{J}}_{\boldsymbol{\eta}}(t) = \Omega_{\mathfrak{p}(t)}\big(\Theta_{\mathfrak{f}}(y), \mathbf{J}_{\boldsymbol{\eta}}(t)\big)\Theta_{\mathfrak{f}}(y) \text{ in } [0,1] \text{ with } \mathbf{J}_{\boldsymbol{\eta}}(0) = \mathbf{0} \;\&\; \mathbf{J}_{\boldsymbol{\eta}}(1) = \boldsymbol{\eta}.$$

Moreover (cf. (9.8)),

$$\mathfrak{E}(\boldsymbol{\eta})_{\mathfrak{F}_{\mathfrak{f}}(y)}\mathrm{Grad}^M U(\,\cdot\,, x) = 2\dot{\mathbf{J}}_{\boldsymbol{\eta}}(1).$$

PROOF: Set

$$\mathfrak{g}_s = \mathrm{Exp}_{\mathfrak{F}_{\mathfrak{f}}(y)}(s\boldsymbol{\eta}) \quad \text{and} \quad \mathfrak{q}_s(t) = \mathrm{Exp}_{\mathfrak{g}_s}\big(t\Theta_{\mathfrak{g}_s}(x)\big)$$

for s near 0 and $t \in [0,1]$. Next, define $\boldsymbol{\xi}_s(t) = \phi\big(\mathfrak{q}_s'(t)\big)$. Clearly $\boldsymbol{\xi}_s(0) = \boldsymbol{\eta}$ and $\boldsymbol{\xi}_s(1) = \mathbf{0}$. In addition, by (8.48) and the fact that $\dot{\mathfrak{q}}_s(t)$ is horizontal,

$$\dot{\boldsymbol{\xi}}_s(t) = \frac{d}{ds}\phi\big(\dot{\mathfrak{q}}_s(t)\big) + \omega\big(\mathfrak{q}_s'(t)\big)\phi\big(\dot{\mathfrak{q}}_s(t)\big).$$

Hence, since $\phi\big(\dot{\mathfrak{q}}_s(t)\big)$ is independent of t, (8.49) yields

$$\ddot{\boldsymbol{\xi}}_s(t) = \Omega_{\mathfrak{q}_s(t)}\Big(\phi\big(\dot{\mathfrak{q}}_s(t)\big), \boldsymbol{\xi}_s(t)\Big)\phi\big(\dot{\mathfrak{q}}_s(t)\big).$$

But clearly $\mathfrak{q}_0(t) = \mathfrak{p}(1-t)$, and therefore we now see that we can take $\mathbf{J}_{\boldsymbol{\eta}}(t) = \boldsymbol{\xi}_0(1-t)$. Moreover, by (8.48),

$$\begin{aligned}\mathfrak{E}(\boldsymbol{\eta})_{\mathfrak{g}}\mathrm{Grad}^M U(\,\cdot\,, x) &= -2\mathfrak{E}(\boldsymbol{\eta})_{\mathfrak{g}}\Theta_{\cdot}(x) = -2\frac{d}{ds}\phi\big(\dot{\mathfrak{q}}_s(0)\big)\Big|_{s=0} \\ &= -2\frac{d}{dt}\phi\big(\mathfrak{q}_0'(t)\big)\Big|_{t=0} + \omega\big(\mathfrak{q}_0'(0)\big)\phi\big(\dot{\mathfrak{q}}_0(0)\big) = -2\dot{\boldsymbol{\xi}}_0(0),\end{aligned}$$

[1] In fact, it was RIemann's analysis of this formnlation which led to his realization that curvature is the obstruction to flatness. See volume II of M. Spivak's great American differential geometry book [34].

since $\mathfrak{q}_0'(0)$ is horizontal. Thus, all that remains is to prove that (9.15) has at most one solution for each $\boldsymbol{\eta}$, and, by linearity, this comes down to showing that the only solution when $\boldsymbol{\eta} = \mathbf{0}$ is identically equal to $\mathbf{0}$. To this end, let $\mathcal{J}$ denote the set of all smooth $\mathbf{J} : [0,1] \longrightarrow \mathbb{R}^d$ satisfying $\mathbf{J}(0) = \mathbf{0}$ and $\ddot{\mathbf{J}}(t) = \Omega_{\mathfrak{p}(t)}\big(\Theta_{\mathfrak{f}}(y), \mathbf{J}(t)\big)\Theta_{\mathfrak{f}}(y)$. Then, because each $\mathbf{J} \in \mathcal{J}$ is uniquely determined by $\dot{\mathbf{J}}(0)$, $\mathcal{J}$ is a d-dimensional vector space. Moreover, we have already shown that the linear map $\mathbf{J} \in \mathcal{J} \longmapsto \mathbf{J}(1) \in \mathbb{R}^d$ is surjective, and so $\mathbf{J} \equiv \mathbf{0}$ can be the only element of its kernel. $\square$

Let $\mathbf{J} : [0,1] \longrightarrow \mathbb{R}^d$ be a solution to the second order differential equation in (9.15), and set $Z(t) = \mathfrak{p}(t)\mathbf{J}(t)$. Then $t \in [0,1] \longmapsto Z(t) \in T_{\pi\circ\mathfrak{p}(t)}M$ is called a *Jacobi field* along the geodesic $\pi \circ \mathfrak{p}$.

9.16 THEOREM. *Given $\mathfrak{f} \in \pi^{-1}x$, $y \in N_x$, and $\boldsymbol{\eta} \in \mathbb{R}^d$ determine $\mathbf{J}_{\boldsymbol{\eta}}$ by (9.15). Then,*

$$\Sigma(\mathfrak{f}, \mathfrak{F}_{\mathfrak{f}}(y))\boldsymbol{\eta} = \dot{\mathbf{J}}_{\boldsymbol{\eta}}(0). \tag{9.17}$$

PROOF: Set

$$y_s = \exp_{\mathfrak{F}_{\mathfrak{f}}(y)}(s\boldsymbol{\eta}) \text{ and } \mathfrak{p}_s(t) = \operatorname{Exp}_{\mathfrak{f}}\big(t\Theta_{\mathfrak{f}}(y_s)\big)$$

for s near 0. Then, by the second equality in (9.7),

$$\begin{aligned}
2\Sigma(\mathfrak{f}, \mathfrak{F}_{\mathfrak{f}}(y))\boldsymbol{\eta} &= \frac{d}{ds}\operatorname{Grad}^M_{\mathfrak{F}_{\mathfrak{f}}(y_s)} U(\,\cdot\,, x)\Big|_{s=0} \\
&= \Big(\mathfrak{E}(\boldsymbol{\eta})_{\mathfrak{F}_{\mathfrak{f}}(y)} + \lambda\big(\omega(\mathfrak{p}_0'(1))\big)_{\mathfrak{F}_{\mathfrak{f}}(y)}\Big)\operatorname{Grad}^M U(\,\cdot\,, x) \\
&= \mathfrak{E}(\boldsymbol{\eta})_{\mathfrak{F}_{\mathfrak{f}}(y)}\operatorname{Grad}^M U(\,\cdot\,, x) - 2\omega(\mathfrak{p}_0'(1))\Theta_{\mathfrak{f}}(y),
\end{aligned}$$

since $\lambda(a)\operatorname{Grad}^M f \circ \pi = -a\operatorname{Grad}^M f \circ \pi$ for any $f \in C^1(M;\mathbb{R})$ and $a \in \mathfrak{o}(\mathbb{R}^d)$.

Next, note that $\mathfrak{p}_0 = \mathfrak{p}$, where $\mathfrak{p}(t) = \operatorname{Exp}_{\mathfrak{f}}\big(t\Theta_{\mathfrak{f}}(y)\big)$. Hence, if $\boldsymbol{\xi}_s(t) \equiv \phi\big(\mathfrak{p}_s'(t)\big)$, then, just as in the first step of the proof of Lemma 9.14,

$$\ddot{\boldsymbol{\xi}}_0(t) = \Omega_{\mathfrak{p}(t)}\big(\Theta_{\mathfrak{f}}(y), \boldsymbol{\xi}_0(t)\big)\Theta_{\mathfrak{f}}(y) \quad \text{with } \boldsymbol{\xi}_0(0) = \mathbf{0} \;\&\; \boldsymbol{\xi}_0(1) = \boldsymbol{\eta}.$$

By the uniqueness statement in Lemma 9.14, this means that $\boldsymbol{\xi}_0 = \mathbf{J}_{\boldsymbol{\eta}}$ and therefore, by that lemma, that $\mathfrak{E}(\boldsymbol{\eta})_{\mathfrak{F}_{\mathfrak{f}}(y)}\operatorname{Grad}^M U(\,\cdot\,, x) = 2\dot{\boldsymbol{\xi}}_0(1)$. At the same time, by (8.49)

$$\omega(\mathfrak{p}_0'(1))\Theta_{\mathfrak{f}}(y) = \int_0^1 \Omega_{\mathfrak{q}_0(t)}\big(\Theta_{\mathfrak{f}}(y), \boldsymbol{\xi}_0(t)\big)\Theta_{\mathfrak{f}}(y)\,dt = \dot{\boldsymbol{\xi}}_0(1) - \dot{\boldsymbol{\xi}}_0(0).$$

Thus, when we combine these and remember that $\boldsymbol{\xi}_0 = \mathbf{J}_{\boldsymbol{\eta}}$, we arrive at (9.17). $\square$

COROLLARY 9.18. *For all* $\mathfrak{f} \in \pi^{-1}x$ *and* $y \in N_x$,

$$\Sigma\big(\mathfrak{f}, \mathfrak{F}_{\mathfrak{f}}(y)\big)\Theta_{\mathfrak{f}}(y) = \Theta_{\mathfrak{f}}(y) = \Sigma\big(\mathfrak{f}, \mathfrak{F}_{\mathfrak{f}}(y)\big)^{\top}\Theta_{\mathfrak{f}}(y),$$

and therefore

$$g^{\Theta_{\mathfrak{f}}}(y)\Theta_{\mathfrak{f}}(y) = \Theta_{\mathfrak{f}}(y).$$

PROOF: First note that $t \in [0,1] \longmapsto t\Theta_{\mathfrak{f}}(y) \in \mathbb{R}^d$ solves (9.15) and is therefore equal to $\mathbf{J}_{\Theta_{\mathfrak{f}}(y)}$. Hence, by (9.17), we have now proved that $\Sigma\big(\mathfrak{f}, \mathfrak{F}_{\mathfrak{f}}(y)\big)\Theta_{\mathfrak{f}}(y) = \Theta_{\mathfrak{f}}(y)$. Next, for any $\boldsymbol{\eta} \in \mathbb{R}^d$,

$$\frac{d^2}{dt^2}\big(\mathbf{J}_{\boldsymbol{\eta}}(t), \Theta_{\mathfrak{f}}(y)\big)_{\mathbb{R}^d} = \Big(\Omega_{\mathfrak{p}(t)}\big(\Theta_{\mathfrak{f}}, \mathbf{J}_{\boldsymbol{\eta}}(t)\big)\Theta_{\mathfrak{f}}(y), \Theta_{\mathfrak{f}}(y)\Big)_{\mathbb{R}^d} = 0.$$

Hence, $\big(\mathbf{J}_{\boldsymbol{\eta}}(t), \Theta_{\mathfrak{f}}(y)\big)_{\mathbb{R}^d} = t\big(\boldsymbol{\eta}, \Theta_{\mathfrak{f}}(y)\big)_{\mathbb{R}^d}$, and so, by (9.17),

$$\Big(\Sigma\big(\mathfrak{f}, \mathfrak{F}_{\mathfrak{f}}(y)\big)\boldsymbol{\eta}, \Theta_{\mathfrak{f}}(y)\Big)_{\mathbb{R}^d} = \big(\dot{\mathbf{J}}_{\boldsymbol{\eta}}(0), \Theta_{\mathfrak{f}}(y)\big)_{\mathbb{R}^d} = \big(\boldsymbol{\eta}, \Theta_{\mathfrak{f}}(y)\big)_{\mathbb{R}^d}.$$

Equivalently, $\Sigma\big(\mathfrak{f}, \mathfrak{F}_{\mathfrak{f}}(y)\big)^{\top}\Theta_{\mathfrak{f}}(y) = \Theta_{\mathfrak{f}}(y)$. Finally, $g^{\Theta_{\mathfrak{f}}}(y)\Theta_{\mathfrak{f}}(y) = \Theta_{\mathfrak{f}}(y)$ now follows from these together with (9.11). □

The second statement in the preceding corollary says that, as one moves along the geodesic ray running from x to y, the $g^{\Theta_{\mathfrak{f}}}$-inner product between $\partial^{\Theta_{\mathfrak{f}}}_{\Theta_{\mathfrak{f}}(y)}$ and $\partial^{\Theta_{\mathfrak{f}}}_{\boldsymbol{\eta}}$ does not change. Equivalently, for any $\boldsymbol{\theta}$, $\boldsymbol{\eta} \in \mathbb{R}^d$,

$$\big\langle (\partial^{\Theta_{\mathfrak{f}}}_{\boldsymbol{\theta}})_{\exp_{\mathfrak{f}}(t\boldsymbol{\theta})}, (\partial^{\Theta_{\mathfrak{f}}}_{\boldsymbol{\eta}})_{\exp_{\mathfrak{f}}(t\boldsymbol{\theta})} \big\rangle = \big(\boldsymbol{\theta}, \boldsymbol{\eta}\big)_{\mathbb{R}^d}$$

as long as $t \rightsquigarrow \exp_{\mathfrak{f}}(t\boldsymbol{\theta})$ stays away from the cut locus. Like so many other key observations, this one is known as *Gauss's Lemma*.

9.19 LEMMA. *Set* $\rho_x(y) \equiv \operatorname{dist}^M(y,x)$, *and, for* $y \in N_x$, *define* $S_{\mathfrak{f}}(y) \in \operatorname{Hom}(\mathbb{R}^d;\mathbb{R}^d)$ *and* $S'_{\mathfrak{f}}(y) \in \operatorname{Hom}(\mathbb{R}^d;\mathbb{R}^d)$ *so that* $S_{\mathfrak{f}}(y) = S'_{\mathfrak{f}}(y) = 0$ *if* $y = x$ *and, when* $y \in N_x \setminus \{x\}$,

$$\begin{aligned} S_{\mathfrak{f}}(y)\boldsymbol{\eta} &= -\Omega_{\mathfrak{f}}\big(\widehat{\Theta_{\mathfrak{f}}}(y), \boldsymbol{\eta}\big)\widehat{\Theta_{\mathfrak{f}}}(y) \\ S'_{\mathfrak{f}}(y)\boldsymbol{\eta} &= -\mathfrak{E}\big(\widehat{\Theta_{\mathfrak{f}}}(y)\big)_{\mathfrak{f}}\Omega\big(\widehat{\Theta_{\mathfrak{f}}}(y), \boldsymbol{\eta}\big)\widehat{\Theta_{\mathfrak{f}}}(y) \end{aligned} \qquad \text{where } \widehat{\Theta_{\mathfrak{f}}}(y) \equiv \rho_x(y)^{-1}\Theta_{\mathfrak{f}}(y).$$

Next, for any $y \in N_x \setminus \{x\}$ *and* $\boldsymbol{\eta} \in \mathbb{S}^{d-1}$, *determine* $\mathbf{J}_{\boldsymbol{\eta}}$ *as in (9.15). Then, there is a* $C \in [1,\infty)$ *such that, for all* $t \in [0,1]$,

$$\left|\dot{\mathbf{J}}_{\boldsymbol{\eta}}(t) - \boldsymbol{\eta} - \rho_x(y)^2\left(\frac{t^2}{2} - \frac{1}{6}\right)S_{\mathfrak{f}}(y)\boldsymbol{\eta} + \rho_x(y)^3\left(\frac{t^3}{3} - \frac{1}{12}\right)S'_{\mathfrak{f}}(y)\boldsymbol{\eta}\right|_{\mathbb{R}^d}$$

is dominated by $C\rho_x(y)^4$ *whenever* $\rho_x(y)^2 \le C^{-1}$.

PROOF: Clearly, we need only handle $y \in N_x \setminus \{x\}$ with $\operatorname{dist}^M(y,x) \le 1$. Given such a y, set $\rho = \rho_x(y)$, $\boldsymbol{\theta} = \Theta_{\mathfrak{f}}(y)$, $S = S_{\mathfrak{f}}(y)$, and $S' = S'_{\mathfrak{f}}(y)$. Because (cf. (8.54) and (R4)) S is symmetric, we can find $(\lambda_1, \dots, \lambda_d) \in \mathbb{R}^d$ and an orthonormal basis $(\mathbf{e}_1, \dots, \mathbf{e}_d)$ in $\mathbb{R}^d$ so that $\lambda_m \mathbf{e}_m = S\mathbf{e}_m$ for $1 \le m \le d$; and so our task comes down to checking that, for each $1 \le m \le d$,

$$(*) \qquad \dot{\mathbf{J}}_m(t) = \mathbf{e}_m - \rho^2\left(\frac{t^2}{2} - \frac{1}{6}\right)\lambda_m \mathbf{e}_m - \rho^3\left(\frac{t^3}{3} - \frac{1}{12}\right)S'\mathbf{e}_m + \mathcal{O}(\rho^4),$$

where $\ddot{\mathbf{J}}_m(t) = \Omega_{\mathrm{Exp}_{\mathfrak{f}}(t\boldsymbol{\theta})}\big(\boldsymbol{\theta}, \mathbf{J}_m(t)\big)\boldsymbol{\theta}$ with $\mathbf{J}_m(0) = \mathbf{0}$ and $\mathbf{J}_m(1) = \mathbf{e}_m$.

In order to prove (*), we have to distinguish between three cases: $\lambda_m = 0$, $\lambda_m > 0$, and $\lambda_m < 0$. When $\lambda_m = 0$, consider the function $t \in [0,1] \longmapsto \boldsymbol{\xi}(t) \in \mathbb{R}^d$ given by

$$\boldsymbol{\xi}(t) = t\mathbf{e}_m - \frac{(t^4 - t)\rho^3}{12} S'\mathbf{e}_m.$$

Then $\boldsymbol{\eta}(t) \equiv \mathbf{J}_m(t) - \boldsymbol{\xi}(t)$ satisfies

$$\begin{aligned}\ddot{\boldsymbol{\eta}}(t) =& \Omega_{\mathrm{Exp}_{\mathfrak{f}}(t\boldsymbol{\theta})}\big(\boldsymbol{\theta}, \boldsymbol{\eta}(t)\big)\boldsymbol{\theta} \\ &+ \Omega_{\mathrm{Exp}_{\mathfrak{f}}(t\boldsymbol{\theta})}\big(\boldsymbol{\theta}, \boldsymbol{\xi}(t)\big)\boldsymbol{\theta} - \Omega_{\mathfrak{f}}\big(\boldsymbol{\theta}, \boldsymbol{\xi}(t)\big)\boldsymbol{\theta} - t\mathfrak{E}(\boldsymbol{\theta})_{\mathfrak{f}}\Omega\big(\boldsymbol{\theta}, \boldsymbol{\xi}(t)\big)\boldsymbol{\theta} + \mathcal{O}(\rho^5),\end{aligned}$$

which means that, for some $C < \infty$,

$$\sup_{t\in[0,1]} \left|\ddot{\boldsymbol{\eta}}(t)\right|_{\mathbb{R}^d} \le \frac{C}{2}\rho^2\left(\rho^2 + \sup_{t\in[0,1]}\left|\boldsymbol{\eta}(t)\right|_{\mathbb{R}^d}\right).$$

Hence, since $\boldsymbol{\eta}(0) = \boldsymbol{\eta}(1) = \mathbf{0}$ and therefore

$$\sup_{t\in[0,1]}\left|\boldsymbol{\eta}(t)\right|_{\mathbb{R}^d} \le \sup_{t\in[0,1]}\left|\dot{\boldsymbol{\eta}}(t)\right|_{\mathbb{R}^d} \le \int_0^1 \left|\ddot{\boldsymbol{\eta}}(t)\right|_{\mathbb{R}^d} dt \le \sup_{t\in[0,1]}\left|\ddot{\boldsymbol{\eta}}(t)\right|_{\mathbb{R}^d},$$

we can conclude, in particular, that $\rho^2 \le C^{-1} \implies |\dot{\boldsymbol{\eta}}(t)|_{\mathbb{R}^d} \le C\rho^4$, which is equivalent to (*). When $\lambda_m \ne 0$, we take $\mu = \rho|\lambda_m|^{\frac{1}{2}}$ and

$$\boldsymbol{\xi}(t) = \alpha(t)\mathbf{e}_m - \rho^3\big(\beta(t) - t\beta(1)\big)S'\mathbf{e}_m,$$

where, depending on whether $\lambda_m < 0$ or $\lambda_m > 0$,

$$\alpha(t) = \frac{\sinh \mu t}{\sinh \mu} \quad \text{or} \quad \alpha(t) = \frac{\sin \mu t}{\sin \mu}$$

and

$$\beta(t) = \frac{\mu t \sinh \mu t + 2(1 - \cosh \mu t)}{\mu^3 \sinh \mu} \quad \text{or} \quad \beta(t) = \frac{-\mu t \sin \mu t + 2(1 - \cos \mu t)}{\mu^3 \sin \mu}.$$

Then, using the facts that $\boldsymbol{\xi}(t) = \mathbf{J}_m(t)$ for $t \in \{0,1\}$, $\ddot{\alpha}(t) = -\rho^2\lambda_m\alpha(t)$, $\ddot{\beta}(t) = t\alpha(t)$, and

$$\dot{\alpha}(t) = 1 - \rho^2\lambda_m\left(\frac{t^2}{2} - \frac{1}{6}\right) + \mathcal{O}(\rho^4) \text{ while } \dot{\beta}(t) = \frac{t^3}{3} - \frac{1}{12} + \mathcal{O}(\rho^4),$$

we can repeat the preceding argument to again arrive at (*). □

By combining Lemma 9.19 with Lemma 9.14 and (9.17), we obtain the expansion alluded to earlier.

9.20 THEOREM. *Refer to the notation introduced in Lemma 9.19. Then (cf. (8.70))*

$$\tfrac{1}{2}\mathrm{Hess}^M_{\mathfrak{F}_\mathfrak{f}(y)}U(\,\cdot\,,x) = I_{\mathbb{R}^d} - \frac{\rho_x(y)^2}{3}S_\mathfrak{f}(y) - \frac{\rho_x(y)^3}{4}S'_\mathfrak{f}(y) + \mathcal{O}\big(\rho_x(y)^4\big),$$

$$\Sigma\big(\mathfrak{f},\mathfrak{F}_\mathfrak{f}(y)\big) = I_{\mathbb{R}^d} + \frac{\rho_x(y)^2}{6}S_\mathfrak{f}(y) + \frac{\rho_x(y)^3}{12}S'_\mathfrak{f}(y) + \mathcal{O}\big(\rho_x(y)^4\big), \tag{9.21}$$

and therefore, by (9.11),

$$g^{\Theta_\mathfrak{f}}(y) = I_{\mathbb{R}^d} - \frac{\rho_x(y)^2}{3}S_\mathfrak{f}(y) - \frac{\rho_x(y)^3}{6}S'_\mathfrak{f}(y) + \mathcal{O}\big(\rho_x(y)^4\big).$$

9.3.2. The Laplacian in Non-Divergence Form. Let $x \in M$ be given, and define the *injection radius* at x to be (cf. (9.3))

$$R_x \equiv \min\big\{R_\mathfrak{f}(\mathbf{e}) : \mathbf{e} \in \mathbf{S}^{d-1}\big\}, \tag{9.22}$$

where $\mathfrak{f}$ is any frame above x. Next, select $\mathfrak{f} \in \pi^{-1}x$ so that (cf. (8.58)) $\mathfrak{R}_\mathfrak{f}$ is diagonal, and define $a(\boldsymbol{\theta}) \in \mathrm{Hom}(\mathbb{R}^d;\mathbb{R}^d)$ and $\mathbf{b}(\boldsymbol{\theta}) \in \mathbb{R}^d$ for $|\boldsymbol{\theta}| < R_x$ so that

$$\begin{aligned} a(\boldsymbol{\theta}) &= \big(\big(a(\boldsymbol{\theta})^{i,j}\big)\big)_{1\le i,j\le d} \equiv \big(g^{\Theta_\mathfrak{f}}\big)^{-1} \circ \exp_\mathfrak{f}(\boldsymbol{\theta}) \\ b^i(\boldsymbol{\theta}) &= \sqrt{\det a(\boldsymbol{\theta})}\sum_{j=1}^d \frac{\partial}{\partial\theta^j}\left(\frac{a^{i,j}}{\sqrt{\det a}}(\boldsymbol{\theta})\right). \end{aligned} \tag{9.23}$$

Then, by (4.25), for any $\varphi \in C^2_c\big(B_{\mathbb{R}^d}(\mathbf{0},R_x);\mathbb{R}\big)$,

$$\Delta_M\big(\varphi\circ\Theta_\mathfrak{f}\big) = \sum_{i,j=1}^d a^{i,j}\frac{\partial^2\varphi}{\partial\theta^i\partial\theta^j}\circ\Theta_\mathfrak{f} + \sum_{i=1}^d b^i\frac{\partial\varphi}{\partial\theta^i}\circ\Theta_\mathfrak{f}. \tag{9.24}$$

9.25 THEOREM. *In (cf. (9.22)) $B_{\mathbb{R}^d}(\mathbf{0};R_x)$,*

$$a^{i,j}(\boldsymbol{\theta}) = \delta^{i,j} - \frac{1}{3}\Big(\Omega_\mathfrak{f}(\boldsymbol{\theta},\mathbf{e}_i)\boldsymbol{\theta},\mathbf{e}_j\Big)_{\mathbb{R}^d} - \frac{1}{6}\mathfrak{C}(\boldsymbol{\theta})_\mathfrak{f}\Big(\Omega(\boldsymbol{\theta},\mathbf{e}_i)\boldsymbol{\theta},\mathbf{e}_j\Big)_{\mathbb{R}^d} + \mathcal{O}\big(|\boldsymbol{\theta}|^4\big)$$

and

$$b^i(\boldsymbol{\theta}) = -\frac{2}{3}\Big(\mathfrak{R}_\mathfrak{f}\boldsymbol{\theta},\mathbf{e}_i\Big)_{\mathbb{R}^d} - \frac{1}{2}\mathfrak{C}(\boldsymbol{\theta})_\mathfrak{f}\Big(\mathfrak{R}\boldsymbol{\theta},\mathbf{e}_i\Big)_{\mathbb{R}^d} + \frac{1}{12}\mathfrak{C}(\mathbf{e}_i)_\mathfrak{f}\Big(\mathfrak{R}\boldsymbol{\theta},\boldsymbol{\theta}\Big)_{\mathbb{R}^d} + \mathcal{O}\big(|\boldsymbol{\theta}|^3\big)$$

as $\boldsymbol{\theta} \to \mathbf{0}$.

PROOF: The expansion for $a(\boldsymbol{\theta})$ is an immediate corollary of the expansion for $g^{\Theta_\mathfrak{f}}$ given at the end of Theorem 9.20. All that one has to do is remember that we are dealing here with the inverse of $g^{\Theta_\mathfrak{f}}$ and be careful about the signs.

To prove the expansion for $\mathbf{b}$, first observe that

$$
\begin{aligned}
b^i(\boldsymbol{\theta}) &= \sum_{j=1}^d \left(\frac{\partial a^{i,j}}{\partial \theta^j} - \frac{1}{2} a^{i,j} \frac{\partial \log(\det a)}{\partial \theta^j}\right)(\boldsymbol{\theta}) \\
(*) \qquad &= \sum_{j=1}^d \left(\frac{\partial a^{i,j}}{\partial \theta^j} - \frac{1}{2}\sum_{k,\ell=1}^d a^{i,j} a_{k,\ell} \frac{\partial a^{k,\ell}}{\partial \theta^j}\right)(\boldsymbol{\theta}) \\
&= \sum_{j=1}^d \frac{\partial a^{i,j}}{\partial \theta^j}(\boldsymbol{\theta}) - \frac{1}{2}\sum_{k=1}^d \frac{\partial a^{k,k}}{\partial \theta^i}(\boldsymbol{\theta}) + \mathcal{O}(|\boldsymbol{\theta}|^3),
\end{aligned}
$$

where $((a_{k,\ell}))$ is the inverse of $((a^{i,j}))$ and we have used Cramer's rule in very much the same way as we did in the derivation of (7.19). Second, note that, for any (i,j):

$$
\begin{aligned}
\frac{\partial a^{i,j}}{\partial \theta^j}(\boldsymbol{\theta}) &= -\frac{1}{3}\Big(\Omega_\mathfrak{f}(\mathbf{e}_j, \mathbf{e}_i)\boldsymbol{\theta}, \mathbf{e}_j\Big)_{\mathbb{R}^d} - \frac{1}{6}\Big(\mathfrak{C}(\mathbf{e}_j)_\mathfrak{f}\Omega(\boldsymbol{\theta}, \mathbf{e}_i)\boldsymbol{\theta}, \mathbf{e}_j\Big)_{\mathbb{R}^d} \\
&\quad - \frac{1}{6}\Big(\mathfrak{C}(\boldsymbol{\theta})_\mathfrak{f}\Omega(\mathbf{e}_j, \mathbf{e}_i)\boldsymbol{\theta}, \mathbf{e}_j\Big)_{\mathbb{R}^d} + \mathcal{O}(|\boldsymbol{\theta}|^3) \\
&= -\frac{1}{3}\Big(\Omega_\mathfrak{f}(\mathbf{e}_j, \mathbf{e}_i)\boldsymbol{\theta}, \mathbf{e}_j\Big)_{\mathbb{R}^d} - \frac{1}{3}\Big(\mathfrak{C}(\boldsymbol{\theta})_\mathfrak{f}\Omega(\mathbf{e}_j, \mathbf{e}_i)\boldsymbol{\theta}, \mathbf{e}_j\Big)_{\mathbb{R}^d} \\
&\quad + \frac{1}{6}\Big(\mathfrak{C}(\mathbf{e}_i)_\mathfrak{f}\Omega(\mathbf{e}_j, \boldsymbol{\theta})\boldsymbol{\theta}, \mathbf{e}_j\Big)_{\mathbb{R}^d} + \mathcal{O}(|\boldsymbol{\theta}|^3),
\end{aligned}
$$

where we have made repeated use of the asymmetry of $\Omega_\mathfrak{f}$ and, in the derivation of the second line, we used (8.52). Hence, by summing over j, we obtain (cf. (8.58))

$$
\begin{aligned}
\sum_{j=1}^d \frac{\partial a^{i,j}}{\partial \theta^j}(\boldsymbol{\theta}) &= -\frac{1}{3}(\mathfrak{R}_\mathfrak{f}\boldsymbol{\theta}, \mathbf{e}_i)_{\mathbb{R}^d} - \frac{1}{3}\mathfrak{C}(\boldsymbol{\theta})_\mathfrak{f}(\mathfrak{R}\,\boldsymbol{\theta}, \mathbf{e}_i)_{\mathbb{R}^d} \\
&\quad + \frac{1}{6}\mathfrak{C}(\mathbf{e}_i)_\mathfrak{f}(\mathfrak{R}\,\boldsymbol{\theta}, \boldsymbol{\theta})_{\mathbb{R}^d} + \mathcal{O}(|\boldsymbol{\theta}|^3).
\end{aligned}
$$

At the same time,

$$
\begin{aligned}
\sum_{k=1}^d \frac{\partial a^{k,k}}{\partial \theta^i}(\boldsymbol{\theta}) &= \frac{1}{3}\frac{\partial}{\partial \theta^i}\left((\mathfrak{R}_\mathfrak{f}\boldsymbol{\theta}, \boldsymbol{\theta})_{\mathbb{R}^d} + \frac{1}{2}\mathfrak{C}(\boldsymbol{\theta})_\mathfrak{f}(\mathfrak{R}\,\boldsymbol{\theta}, \boldsymbol{\theta})_{\mathbb{R}^d}\right) + \mathcal{O}(|\boldsymbol{\theta}|^3) \\
&= \frac{2}{3}(\mathfrak{R}_\mathfrak{f}\boldsymbol{\theta}, \mathbf{e}_i)_{\mathbb{R}^d} + \frac{1}{3}\mathfrak{C}(\boldsymbol{\theta})_\mathfrak{f}(\mathfrak{R}\,\boldsymbol{\theta}, \mathbf{e}_i)_{\mathbb{R}^d} + \frac{1}{6}\mathfrak{C}(\mathbf{e}_i)(\mathfrak{R}\,\boldsymbol{\theta}, \boldsymbol{\theta})_{\mathbb{R}^d} + \mathcal{O}(|\boldsymbol{\theta}|^3).
\end{aligned}
$$

Finally, by plugging these into $(*)$, we arrive at the required expansion. □

9.4 Coupling

The calculations in Theorems 9.20 and 9.25 should make one believe that, up to the time when it leaves a small ball around x, the Brownian motion on M starting at x should be well approximated by the composition of $\exp_{\mathfrak{f}}$ with a Euclidean Brownian motion.[2] Equivalently, if $\mathbf{w}(t,p) \equiv \Theta_{\mathfrak{f}} \circ p(t)$, $p \in \mathcal{P}(M)$, then, for $\mathbb{P}_x^M$-typical p's, we should expect that, until it escapes from a small ball around the origin, $\mathbf{w}(\,\cdot\,,p)$ will behave very much like a typical Euclidean Brownian path. The purpose of this subsection is to provide this expectation with substance. To be more precise, we will prove the following *coupling*[3] result.

9.26 THEOREM. *For $r > 0$, define $\tau_r : \mathfrak{W}(\mathbb{R}^d) \longrightarrow [0,\infty]$ by*

$$\tau_r(\mathbf{w}) = \inf\{t \geq 0 : |\mathbf{w}(t)|_{\mathbb{R}^d} \geq r\},$$

and set (cf. (9.22)) $r_x = \frac{R_x}{2} \wedge 1$. Then there exists a Borel probability measure $\mathbb{Q}$ on $\mathfrak{W}(\mathbb{R}^d)^2$ with the properties that, when $\mathbb{Q}^{(i)}$, $i \in \{1,2\}$, denotes the $\mathbb{Q}$-distribution of $(\mathbf{w}_1,\mathbf{w}_2)\rightsquigarrow\mathbf{w}_i$:

$$\mathbb{Q}^{(2)} = \mu_{\mathbb{R}^d}, \quad \mathbb{Q}^{(1)} \restriction \mathcal{B}_{\tau_{r_x}} = (\Theta_{\mathfrak{f}})_* \mathbb{P}_x^M \restriction \mathcal{B}_{\tau_{r_x}},$$

and, for some $K \in [1,\infty)$ and $r_0 \in (0,r_x]$,

$$\mathbb{Q}\left(\sup_{0 \leq t < \tau_r(\mathbf{w}_1)} \left|\mathbf{w}_1(t) - \mathbf{w}_2(t)\right|_{\mathbb{R}^d} \geq Lr^3\right) \leq Ke^{-\frac{L}{K}}, \quad r \in (0,r_0] \text{ and } L > 0.$$

In order to construct such a $\mathbb{Q}$, first choose $\eta \in C_c^\infty\big(B_{\mathbb{R}^d}(\mathbf{0},R_x);[0,1]\big)$ so that $\eta \equiv 1$ on $\overline{B_{\mathbb{R}^d}(\mathbf{0},r_x)}$. Now, define $(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \in \mathbb{R}^d \times \mathbb{R}^d \longmapsto \boldsymbol{\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \in \mathbb{R}^d \times \mathbb{R}^d$ and $(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \in \mathbb{R}^d \times \mathbb{R}^d \longmapsto \sigma(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \in \mathrm{Hom}(\mathbb{R}^d;\mathbb{R}^d \times \mathbb{R}^d)$ so that (cf. (9.23))

$$\boldsymbol{\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \frac{1}{2}\begin{pmatrix} (\eta\mathbf{b})(\boldsymbol{\theta}_1) \\ \mathbf{0} \end{pmatrix} \quad \text{or} \quad \boldsymbol{\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$

and (cf. (9.9))

$$\sigma(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \begin{pmatrix} \eta(\boldsymbol{\theta}_1)\Sigma\big(\mathfrak{f},\exp_{\mathfrak{f}}(\boldsymbol{\theta}_1)\big) \\ I_{\mathbb{R}^d} \end{pmatrix} \quad \text{or} \quad \sigma(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \begin{pmatrix} 0 \\ I_{\mathbb{R}^d} \end{pmatrix}$$

[2] This is the observation on which R. Gangoli [18] based his construction of Brownian motion on a Riemannian manifold. His construction was one of the first which made essential use of the Riemannian structure.

[3] For the benefit of the uninitiated, suffice it to say that a "coupling of random variables" X_1 and X_2 defined on probability spaces $(\Omega_1,\mathcal{B}_1,\mathbb{P}_1)$ and $(\Omega_2,\mathcal{B}_2,\mathbb{P}_2)$ and taking values in measurable spaces $(E_1,\mathcal{F}_1)$ and $(E_2,\mathcal{F}_2)$ is any probability measure $\mathbb{Q}$ on $(E_1 \times E_2,\mathcal{F}_1 \times \mathcal{F}_2)$ with the property the $\mathbb{Q}$-distribution of the ith coordinate coincides with the $\mathbb{P}_i$-distribution of X_i for $i \in \{1,2\}$. The purpose of a coupling is to provide a context in which it is possible to make a direct comparison of the random variables involved.

according to whether $|\boldsymbol{\theta}_1|_{\mathbb{R}^d} < R_x$ or $|\boldsymbol{\theta}_1|_{\mathbb{R}^d} \geq R_x$. Next, define the operator $\mathcal{L}$ on $C^2(\mathbb{R}^d \times \mathbb{R}^d; \mathbb{R})$ so that

$$\mathcal{L}\Phi = \tfrac{1}{2}\mathrm{Trace}\big(\sigma^\top (D^2\Phi)\sigma\big) + \big(\boldsymbol{\beta}, D\Phi\big)_{\mathbb{R}^d},$$

where $D\Phi$ and $D^2\Phi$ denote, respectively, the standard Euclidean gradient and Hessian of Φ. Finally, let $\mathbb{Q}$ denote the solution to the martingale problem for $\mathcal{L}$ starting at $(\mathbf{0},\mathbf{0})$.

Obviously, $\mathbb{Q}^{(2)}$ solves the martingale problem for $\frac{1}{2}\Delta_{\mathbb{R}^d}$ starting at $\mathbf{0}$, and so $\mathbb{Q}^{(2)} = \mu_{\mathbb{R}^d}$. In addition, $\mathbb{Q}^{(1)}$ solves the martingale problem for (cf. (9.23))

$$\mathcal{L}^{(1)} \equiv \frac{1}{2}\left(\sum_{i,j=1}^{d} \eta^2 a^{i,j}\frac{\partial^2}{\partial\theta^i\partial\theta^j} + \sum_{i=1}^{d}\eta b^i \frac{\partial}{\partial\theta^i}\right)$$

starting at $\mathbf{0}$. Hence, since (cf. (9.24))

$$\mathcal{L}^{(1)}\varphi = \Delta_M(\varphi\circ\Theta_{\mathfrak{f}}) \quad \text{for } \varphi \in C^2\big(B_{\mathbb{R}^d}(\mathbf{0}, r_x);\mathbb{R}\big),$$

Corollary 3.11 implies that $\mathbb{Q}^{(1)}$ coincides with $(\Theta_{\mathfrak{f}})_*\mathbb{P}_x^M$ on $\mathcal{B}_{\tau_{r_x}}$. Thus, all that remains is to check that the joint distribution of $\mathbf{w}_1$ and $\mathbf{w}_2$ under $\mathbb{Q}$ satisfies the required estimate.

9.27 LEMMA. *There exists a $K \in [1,\infty)$ such that, for each $0 < r \leq r_x$ and $T > 0$ and all $R > 0$,*

$$\mathbb{Q}\left(\sup_{0\leq t\leq T\wedge\tau_r(\mathbf{w}_1)}\left|\mathbf{w}_1(t) - \tfrac{1}{2}\int_0^t (\eta\mathbf{b})\big(\mathbf{w}_1(\tau)\big)\,d\tau - \mathbf{w}_2(t)\right| \geq R\right) \leq K\exp\left(-\frac{R^2}{Kr^4T}\right).$$

PROOF: Set

$$\mathbf{Y}(t,\mathbf{w}) = \mathbf{w}_1(t) - \tfrac{1}{2}\int_0^t (\eta\mathbf{b})\big(\mathbf{w}_1(\tau)\big)\,d\tau - \mathbf{w}_2(t)$$

and

$$B(t,\mathbf{w}) = \eta\big(\mathbf{w}_1(t)\big)\Sigma\Big(\mathfrak{f}, \exp_{\mathfrak{f}}\big(\mathbf{w}_1(t)\big)\Big)^\top - I_{\mathbb{R}^d}$$

for $\mathbf{w} = (\mathbf{w}_1, \mathbf{w}_2)$. Then, by the first part of Theorem 2.42 and Doob's Stopping Time Theorem, for each $r > 0$ and $\boldsymbol{\theta} \in \mathbb{R}^d$,

$$M_r^{\boldsymbol{\theta}}(t,\mathbf{w}) = \exp\left(\Big(\boldsymbol{\theta}, \mathbf{Y}\big(t\wedge\tau_r(\mathbf{w}_1),\mathbf{w}\big)\Big)_{\mathbb{R}^d} - \frac{1}{2}\int_0^{t\wedge\tau_r(\mathbf{w}_1)} \big|B(\tau,\mathbf{w})\boldsymbol{\theta}\big|_{\mathbb{R}^d}^2\,d\tau\right)$$

is a $\mathbb{Q}$-martingale. Moreover, by Theorem 9.20, there is a $C<\infty$ such that

$$\left|B(\tau,\mathbf{w})\boldsymbol{\theta}\right|_{\mathbb{R}^d}^2 \le Cr^4|\boldsymbol{\theta}|^2, \quad \boldsymbol{\theta}\in\mathbb{R}^d \text{ and } 0\le\tau\le\tau_r(\mathbf{w}_1)$$

for all $0<r\le r_x$. Hence, by the version of Doob's Inequality in (7.1.9) of [35], for any $\boldsymbol{\theta}\in\mathbb{R}^d\setminus\{\mathbf{0}\}$,

$$\mathbb{Q}\left(\sup_{0\le t\le T\wedge\tau_r(\mathbf{w}_1)} \big(\hat{\boldsymbol{\theta}},\mathbf{Y}(t,\mathbf{w})\big)_{\mathbb{R}^d}\big)\ge R\right)\le \exp\Big(-R|\boldsymbol{\theta}|_{\mathbb{R}^d}+\tfrac12 Cr^4T|\boldsymbol{\theta}|_{\mathbb{R}^d}^2\Big),$$

where $\hat{\boldsymbol{\theta}}=\frac{\boldsymbol{\theta}}{|\boldsymbol{\theta}|}$. After minimizing over $|\boldsymbol{\theta}|$ for a fixed $\hat{\boldsymbol{\theta}}\in\mathbf{S}^{d-1}$, we conclude that

$$\mathbb{Q}\left(\sup_{0\le t\le T\wedge\tau_r(\mathbf{w}_1)} \big(\hat{\boldsymbol{\theta}},\mathbf{Y}(t,\mathbf{w})\big)_{\mathbb{R}^d}\big)\ge R\right)\le \exp\left(-\frac{R^2}{2Cr^4T}\right).$$

Finally, by applying the preceeding when $\hat{\boldsymbol{\theta}}=\pm\mathbf{e}_k$ for $1\le k\le d$, we arrive at

$$\mathbb{Q}\left(\sup_{0\le t\le T\wedge\tau_r(\mathbf{w}_1)} \big|\mathbf{Y}(t,\mathbf{w})\big|_{\mathbb{R}^d}\ge R\right)\le 2d\exp\left(-\frac{R^2}{2dCr^4T}\right). \quad \square$$

The estimate in Lemma 9.27 brings us most of the way. Indeed, all that remains is to show that there exists a $K\in[1,\infty)$ and an $r_0\in(0,r_x]$ such that

$$\text{(9.28)} \qquad \mathbb{Q}\big(\tau_r(\mathbf{w}_1)\ge T\big)\le Ke^{-\frac{T}{Kr^2}}, \quad 0<r\le r_0 \text{ and } T>0.$$

To prove this, set

$$\psi_r(t,\boldsymbol{\theta})=\exp\left(\frac{\pi^2 t}{32r^2}\right)\cos\frac{\pi\theta^1}{4r}, \quad (t,\boldsymbol{\theta})\in[0,\infty)\times\mathbb{R}^d,$$

and observe that, by Theorem 9.25,

$$\left(\frac{\partial\psi_r}{\partial t}+\mathcal{L}^{(1)}\psi_r\right)(t,\boldsymbol{\theta})=-\frac{\pi^2}{32r^2}\big(1-\mathcal{O}(r^2)\big)\psi_r(t,\theta)$$

for $(t,\boldsymbol{\theta})\in[0,\infty)\times B_{\mathbb{R}^d}(\mathbf{0},r)$. Hence, we can choose $r_0\in(0,r_x]$ so that

$$\left(\frac{\partial\psi_r}{\partial t}+\mathcal{L}^{(1)}\psi_r\right)(t,\boldsymbol{\theta})\le 0 \quad \text{for all } (t,\boldsymbol{\theta})\in[0,\infty)\times B_{\mathbb{R}^d}(\mathbf{0},r)$$

and $0<r\le r_0$. In particular, because

$$\psi_r\big(t,\mathbf{w}_1(t)\big)-\int_0^t\left(\frac{\partial\psi_r}{\partial t}+\mathcal{L}^{(1)}\psi_r\right)\big(t,\mathbf{w}_1(\tau)\big)\,d\tau$$

is a $\mathbb{Q}$-martingale, Doob's Stopping Time Theorem tells us that

$$\begin{aligned}2^{-\frac12}\exp\left(\frac{\pi^2T}{32r^2}\right)&\mathbb{Q}\big(\tau_r(\mathbf{w}_1)\ge T\big)\\ &\le\mathbb{E}^{\mathbb{Q}}\Big[\psi_r\big(T\wedge\tau_r(\mathbf{w}_1),\mathbf{w}_1(T\wedge\tau_r)\big)\Big]\le\psi_r(0,\mathbf{0})=1\end{aligned}$$

for all $0<r\le r_0$ and $T>0$, and clearly (9.28) is a trivial consequence of this.

PROOF OF THEOREM 9.26: Let r_0 be as in (9.28), and use Theorem 9.25 to choose $C < \infty$ so that $|\mathbf{b}(\boldsymbol{\theta})|_{\mathbb{R}^d} \le C|\boldsymbol{\theta}|_{\mathbb{R}^d}$ for $|\boldsymbol{\theta}|_{\mathbb{R}^d} \le r_0$. Then, for any $0 < r \le r_0$, $R > 0$, and $T > 0$,

$$\begin{aligned}
&\mathbb{Q}\left(\sup_{0\le t<\tau_r(\mathbf{w}_1)} \left|\mathbf{w}_1(t) - \mathbf{w}_2(t)\right|_{\mathbb{R}^d} \ge 2R\right) \\
&\le \mathbb{Q}\left(\sup_{0\le t\le T\wedge\tau_r(\mathbf{w}_1)} \left|\mathbf{w}_1(t) - \tfrac{1}{2}\int_0^t (\bar{\eta}\mathbf{b})\big(\mathbf{w}_1(\tau)\big)\,d\tau - \mathbf{w}_2(t)\right|_{\mathbb{R}^d} \ge R\right) \\
&\qquad + \mathbb{Q}\left(\int_0^{\tau_r(\mathbf{w}_1)} \left|\mathbf{b}\big(\mathbf{w}_1(t)\big)\right|_{\mathbb{R}^d} dt \ge R\right) + \mathbb{Q}\big(\tau_r(\mathbf{w}_1) > T\big) \\
&\le Ke^{-\frac{R^2}{Kr^4T}} + Ke^{-\frac{R}{CKr^3}} + Ke^{-\frac{T}{Kr^2}}
\end{aligned}$$

where, in the passage to the last line, we have used the estimates in Lemma 9.27 and (9.28) together with the fact that $|\mathbf{b}(\boldsymbol{\theta})|_{\mathbb{R}^d} \le C|\boldsymbol{\theta}|_{\mathbb{R}^d}$. In particular, if $0 < r < r_0$ and we take $T = \frac{R}{r}$, then we obtain

$$\mathbb{Q}\left(\sup_{0\le t<\tau_r(\mathbf{w}_1)} \left|\mathbf{w}_1(t) - \mathbf{w}_2(t)\right|_{\mathbb{R}^d} \ge 2R\right) \le 2Ke^{-\frac{R}{Kr^3}} + Ke^{-\frac{R}{CKr^3}}.$$

Obviously, the required estimate follows after taking $R = Lr^3$ and making minor adjustments in the choice of K. $\square$

Our applications of Theorem 9.27 will be based on the following corollary, in which

$$\zeta_r(p) \equiv \inf\left\{t \ge 0 : \operatorname{dist}^M\big(p(t), p(0)\big) \ge r\right\} \quad \text{for } r > 0 \text{ and } p \in \mathcal{P}(M).$$

9.29 COROLLARY. *There exists an $r_0 > 0$ and $K < \infty$ such that, for all $\psi \in C(\mathbb{R}^d;\mathbb{R})$ satisfying $\psi(0) = 0$ and*

$$\sup_{\boldsymbol{\theta},\boldsymbol{\theta}'\in B_{\mathbb{R}^d}(\mathbf{0},r),\,\boldsymbol{\theta}\ne\boldsymbol{\theta}'} \frac{\left|\psi(\boldsymbol{\theta}') - \psi(\boldsymbol{\theta})\right|}{|\boldsymbol{\theta}' - \boldsymbol{\theta}|_{\mathbb{R}^d}} \le Cr^{m-1} \quad \text{for } r \in (0, r_0]$$

for some $C < \infty$ and $m \ge 1$,

$$\left|\mathbb{E}^{\mathbb{P}_x^M}\left[\int_0^{\zeta_r(p)} \psi\circ\Theta_{\mathfrak{f}}\big(p(t)\big)\,dt\right] - r^2\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\int_0^{\tau_1(\mathbf{w})} \psi\big(r\mathbf{w}(t)\big)\,dt\right]\right|$$

is dominated by $KCr^{m+4}\log\frac{1}{r}$ whenever $0 < r \le r_0$.

PROOF: First, remember (cf. § 1.2.3) that $\mu_{\mathbb{R}^d}$ is invariant under the scaling transformation $\mathbf{w}\rightsquigarrow\mathbf{w}_r$, where $\mathbf{w}_r(t) = r\mathbf{w}(r^{-2}t)$, $t \ge 0$. Thus, since $\tau_r(\mathbf{w}_r) = r^2\tau_1(\mathbf{w})$, we see that

$$\int_0^{\tau_r(\mathbf{w})} \psi\big(\mathbf{w}(t)\big)\,dt \quad \text{and} \quad r^2\int_0^{\tau_1(\mathbf{w})} \psi\big(r\mathbf{w}(t)\big)\,dt$$

have the same $\mu_{\mathbb{R}^d}$-distribution. In particular, we now know that, for $0 < r \le r_x$,

$$\begin{aligned}
&\left|\mathbb{E}^{\mathbb{P}_x^M}\left[\int_0^{\zeta_r(p)} \psi\circ\Theta_{\mathfrak{f}}\big(p(t)\big)\,dt\right] - r^2\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\int_0^{\tau_1(\mathbf{w})}\psi\big(r\mathbf{w}(t)\big)\,dt\right]\right| \\
&\qquad = \left|\mathbb{E}^{\mathbb{Q}}\left[\int_0^{\tau_r(\mathbf{w}_1)}\psi\big(\mathbf{w}_1(t)\big)\,dt - \int_0^{\tau_r(\mathbf{w}_2)}\psi\big(\mathbf{w}_2(t)\big)\,dt\right]\right| \\
&\qquad\quad \le \mathbb{E}^{\mathbb{Q}}\left[\left|\int_0^{\tau_r(\mathbf{w}_1)}\psi\big(\mathbf{w}_1(t)\big)\,dt - \int_0^{\tau_r(\mathbf{w}_2)}\psi\big(\mathbf{w}_2(t)\big)\,dt\right|\right].
\end{aligned}$$

Next, choose r_0 and K be as in Theorem 9.27, let $0 < r \le r_0$ be given, set

$$A = \left\{(\mathbf{w}_1,\mathbf{w}_2) : \sup_{0\le t<\tau_r(\mathbf{w}_1)}\big|\mathbf{w}_1(t)-\mathbf{w}_2(t)\big|_{\mathbb{R}^d} \le 4Kr^3\log\tfrac{1}{r}\right\},$$

and note that

$$(\mathbf{w}_1,\mathbf{w}_2)\in A \implies \big|\tau_r(\mathbf{w}_1)-\tau_r(\mathbf{w}_2)\big| \le \tau_{r_+}(\mathbf{w}_2)-\tau_{r_-}(\mathbf{w}_2),$$

where $r_\pm = r \pm 4Kr^3\log\frac{1}{r}$. Also, define

$$\bar\psi(\boldsymbol{\theta}) = \begin{cases} \psi(\boldsymbol{\theta})\wedge(Cr^m) & \text{if } \psi(\boldsymbol{\theta})\ge 0 \\ \psi(\boldsymbol{\theta})\vee(-Cr^m) & \text{if } \psi(\boldsymbol{\theta})\le 0\end{cases}.$$

Then,

$$\begin{aligned}
&\mathbb{E}^{\mathbb{Q}}\left[\left|\int_0^{\tau_r(\mathbf{w}_1)}\psi\big(\mathbf{w}_1(t)\big)\,dt - \int_0^{\tau_r(\mathbf{w}_2)}\psi\big(\mathbf{w}_2(t)\big)\,dt\right|\right] \\
&\quad = \mathbb{E}^{\mathbb{Q}}\left[\left|\int_0^{\tau_r(\mathbf{w}_1)}\bar\psi\big(\mathbf{w}_1(t)\big)\,dt - \int_0^{\tau_r(\mathbf{w}_2)}\bar\psi\big(\mathbf{w}_2(t)\big)\,dt\right|\right] \\
&\quad \le Cr^m\mathbb{E}^{\mathbb{Q}}\big[\tau_r(\mathbf{w}_1)+\tau_r(\mathbf{w}_2),\, A\complement\big] \\
&\qquad\qquad + \mathbb{E}^{\mathbb{Q}}\left[\int_0^{\tau_r(\mathbf{w}_1)}\big|\bar\psi\big(\mathbf{w}_1(t)\big)-\bar\psi\big(\mathbf{w}_2(t)\big)\big|\,dt,\, A\right] \\
&\qquad\qquad\qquad + \mathbb{E}^{\mathbb{Q}}\left[\left|\int_{\tau_r(\mathbf{w}_1)}^{\tau_r(\mathbf{w}_2)}\bar\psi\big(\mathbf{w}_2(t)\big)\,dt\right|,\, A\right] \\
&\quad \le Cr^m\mathbb{E}^{\mathbb{Q}}\left[\big(\tau_r(\mathbf{w}_1)+\tau_r(\mathbf{w}_2)\big)^2\right]^{\frac12}\mathbb{Q}(A\complement)^{\frac12} \\
&\qquad\qquad + 4CKr^{m+2}\log\tfrac{1}{r}\,\mathbb{E}^{\mathbb{Q}}\big[\tau_r(\mathbf{w}_1)\big] + Cr^m\mathbb{E}^{\mu_{\mathbb{R}^d}}\big[\tau_{r_+}(\mathbf{w})-\tau_{r_-}(\mathbf{w})\big],
\end{aligned}$$

In order to complete the proof from here, we need a couple of simple observations. First, as a consequence of (9.28), note that

$$\mathbb{E}^{\mathbb{Q}}\big[\tau_r(\mathbf{w}_1)\big] \le \sqrt{\mathbb{E}^{\mathbb{Q}}\big[\tau_r(\mathbf{w}_1)^2\big]} \le Br^2$$

for some $B < \infty$. Second, observe that, by the same argument as we used to obtain (9.28), we can show that there is a $K_0 \in [1, \infty)$ such that

$$\mu_{\mathbb{R}^d}\big(\tau_r(\mathbf{w}) \geq T\big) \leq K_0 e^{-\frac{T}{K_0 r^2}} \quad \text{for all } r > 0 \text{ and } T > 0 \tag{9.30}$$

and therefore that

$$\mathbb{E}^{\mathbb{Q}}\big[\tau_r(\mathbf{w}_2)\big] \leq \sqrt{\mathbb{E}^{\mathbb{Q}}\big[\tau_r(\mathbf{w}_2)^2\big]} \leq B_0 r^2$$

for some $B_0 < \infty$. Hence, the first term is dominated by

$$C(B + B_0) r^{m+2} e^{\frac{4K \log r}{2K}} = C(B + B_0) r^{m+4},$$

while the second term is dominated by $4KCBr^{m+4} \log \frac{1}{r}$. Finally, to handle the third term, we need to know that

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\big[\tau_r(\mathbf{w})\big] = \frac{r^2}{d}. \tag{9.31}$$

To see this, observe that $d^{-1}|\mathbf{w}(t)|^2_{\mathbb{R}^d} - t$ is a $\mu_{\mathbb{R}^d}$-martingale, and use Doob's Stopping Time Theorem to conclude that

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\big[\tau_r(\mathbf{w})\big] = \lim_{t \to \infty} \mathbb{E}^{\mu_{\mathbb{R}^d}}\big[t \wedge \tau_r(\mathbf{w})\big] = d^{-1} \lim_{t \to \infty} \mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\mathbf{w}\big(t \wedge \tau_r\big)\Big] = \frac{r^2}{d}.$$

Given (9.31), we see that the third term is dominated by

$$Cr^m\big(r_+^2 - r_-^2\big) = \frac{8KC}{d} r^{m+4} \log \tfrac{1}{r}. \quad \square$$

9.4.1. Applications. In this concluding subsection, we give a couple of examples (see (9.32) and (9.34) below) of the way in which the preceding can be applied. The results which we will give are primitive by comparison to those found by Mark Pinsky and his co-authors (cf. [32] for a concise introduction to their work as well as references to the original papers).

As our first example, we will show that[4]

$$\mathbb{E}^{\mathbb{P}^M_x}\big[\zeta_r(p)\big] = \frac{r^2}{d} + \frac{r^4}{6d^2(d+2)} \mathrm{Trace}(\mathfrak{R}_{\mathfrak{f}}) + \mathcal{O}\big(r^5 \log \tfrac{1}{r}\big). \tag{9.32}$$

[4] The appearance of the logarithmic factor in the error term is an annoying feature of the procedure which we have adopted. It should not be there! Indead, standard perturbation techniques show that the quantity on the left is a smooth function of small $r > 0$. Hence, the fact that the error term is $o(r^4)$ means that it must be $O(r^5)$. However, our methods do not prove this on their own.

To prove (9.32), first note that, by Doob's Stopping Time Theorem,

$$r^2 = \mathbb{E}^{\mathbb{P}_x^M}\left[U\big(\mathfrak{f}, p(\zeta_r)\big)\right] = \mathbb{E}^{\mathbb{P}_x^M}\left[\int_0^{\zeta_r(p)} \tfrac{1}{2}\left[\Delta_M U(\mathfrak{f}, \cdot\,)\right]\big(p(t)\big)\, dt\right],$$

where $\mathfrak{f} \in \pi^{-1}x$ is chosen so that $\mathfrak{R}_\mathfrak{f}$ is diagonal. Next, use either Theorem 9.20 or Theorem 9.25 combined with (9.24) to check that

$$\tfrac{1}{2}\left[\Delta_M U(\mathfrak{f}, \cdot\,)\right](x) = d - \tfrac{1}{3}\big(\mathfrak{R}_\mathfrak{f}\Theta_\mathfrak{f}, \Theta_\mathfrak{f}\big)_{\mathbb{R}^d} - \tfrac{1}{4}\mathfrak{C}(\Theta_\mathfrak{f})_\mathfrak{f}\big(\mathfrak{R}\,\Theta_\mathfrak{f}, \Theta_\mathfrak{f}\big)_{\mathbb{R}^d} + \mathcal{O}(r^4).$$

Hence, by combining these two and applying Corollary 9.29, we arrive at

$$\begin{aligned} d\mathbb{E}^{\mathbb{P}_x^M}\left[\zeta_r(p)\right] = \mathcal{O}\big(r^5 \log\tfrac{1}{r}\big) + r^2 + \frac{r^4}{3}\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\int_0^{\tau_1(\mathbf{w})} \Big(\mathfrak{R}_\mathfrak{f}\mathbf{w}(t), \mathbf{w}(t)\Big)_{\mathbb{R}^d}\, dt\right] \\ + \frac{r^5}{4}\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\int_0^{\tau_1(\mathbf{w})} \mathfrak{C}\big(\mathbf{w}(t)\big)_\mathfrak{f}\Big(\mathfrak{R}\,\mathbf{w}(t), \mathbf{w}(t)\Big)_{\mathbb{R}^d}\, dt\right]. \end{aligned}$$

Finally, note that, because $-\mathbf{w}$ has the same $\mu_{\mathbb{R}^d}$-distribution as $\mathbf{w}$, the coefficient of r^5 vanishes. At the same time, because we have chosen $\mathfrak{f}$ so that $\mathfrak{R}_\mathfrak{f}$ is diagonal, the coefficient of $\frac{r^4}{3}$ is equal

$$\sum_{i=1}^d \lambda_i \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\int_0^{\tau_1(\mathbf{w})} \big(\mathbf{e}_i, \mathbf{w}(t)\big)_{\mathbb{R}^d}^2\, dt\right],$$

where the λ_i's are the eigenvalues of $\mathfrak{R}_\mathfrak{f}$. But, by the rotation invariance of $\mu_{\mathbb{R}^d}$, it is clear that

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\int_0^{\tau_1(\mathbf{w})} \big(\mathbf{e}_i, \mathbf{w}(t)\big)_{\mathbb{R}^d}^2\, dt\right] = \frac{1}{d}\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\int_0^{\tau_1(\mathbf{w})} |\mathbf{w}(t)|_{\mathbb{R}^d}^2\, dt\right],$$

and, by Doob's Stopping Time Theorem,

$$1 = \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[|\mathbf{w}(\tau_1)|_{\mathbb{R}^d}^4\right] = (2d+4)\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\int_0^{\tau_1(\mathbf{w})} |\mathbf{w}(t)|_{\mathbb{R}^d}^2\, dt\right],$$

and so

$$\text{(9.33)} \qquad \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\int_0^{\tau_1(\mathbf{w})} \big(\mathbf{w}(t), \mathbf{e}_i\big)_{\mathbb{R}^d}^2\, dt\right] = \frac{1}{2d(d+2)}.$$

The trace of the Ricci curvature is called the *mean curvature*, and clearly (9.32) is saying that positive mean curvature causes Brownian motion to spend more time getting out of small balls whereas negative mean curvature has the opposite effect.

We turn next to the $\mathbb{P}_x^M$-distribution of the first exit place $p(\zeta_r)$ of p from $B_M(\mathbf{0},r)$. For this purpose, let $f \in C^3(M;\mathbb{R})$ be given, set $\psi = f \circ \exp_{\mathfrak{f}}$, and define

$$\psi_r(\boldsymbol{\theta}) = \int_{\mathbf{S}^{d-1}} P(\boldsymbol{\theta},\boldsymbol{\eta})\psi(r\boldsymbol{\eta})\,\bar{\lambda}_{\mathbf{S}^{d-1}}(d\boldsymbol{\eta}), \quad \boldsymbol{\theta} \in B_{\mathbb{R}^d}(\mathbf{0},1),$$

where $\bar{\lambda}_{\mathbf{S}^{d-1}}$ is the normalized surface measure on $\mathbf{S}^{d-1}$ and

$$P(\boldsymbol{\theta},\boldsymbol{\eta}) \equiv \frac{1-|\boldsymbol{\theta}|^2_{\mathbb{R}^d}}{|\boldsymbol{\theta}-\boldsymbol{\eta}|^d_{\mathbb{R}^d}}$$

is the Poisson kernel for the unit ball $B_{\mathbb{R}^d}(\mathbf{0},1)$. Then $\boldsymbol{\theta} \in B_{\mathbb{R}^d}(0,r) \longmapsto \psi_r\left(\frac{\boldsymbol{\theta}}{r}\right) \in \mathbb{R}$ is the unique $u \in C\big(\overline{B_{\mathbb{R}^d}(\mathbf{0},r)};\mathbb{R}\big)$ which is $\Delta_{\mathbb{R}^d}$-harmonic in $B_{\mathbb{R}^d}(\mathbf{0},r)$ and equal to ψ on $\partial B_{\mathbb{R}^d}(\mathbf{0},r)$. In particular, $\Delta_{\mathbb{R}^d}\psi_r = 0$,

$$\mathbb{E}^{\mathbb{P}_x^M}\left[f\big(p(\zeta_r)\big)\right] = \mathbb{E}^{\mathbb{P}_x^M}\left[\psi_r\Big(r^{-1}\Theta_{\mathfrak{f}}\big(p(\zeta_r)\big)\Big)\right],$$

and so

$$\begin{aligned}
&\mathbb{E}^{\mathbb{P}_x^M}\left[f\big(p(\zeta_r)\big)\right] - \psi_r(\mathbf{0}) \\
&= \frac{1}{2r^2}\mathbb{E}^{\mathbb{P}_x^M}\left[\int_0^{\zeta_r(p)} \mathrm{Trace}\big((a-I_{\mathbb{R}^d})D^2\psi_r\big)\Big(r^{-1}\Theta_{\mathfrak{f}}\big(p(t)\big)\Big)\,dt\right] \\
&\qquad + \frac{1}{2r}\mathbb{E}^{\mathbb{P}_x^M}\left[\int_0^{\zeta_r(p)} \big(\mathbf{b},D\psi_r\big)_{\mathbb{R}^d}\Big(r^{-1}\Theta_{\mathfrak{f}}\big(p(t)\big)\Big)\,dt\right],
\end{aligned}$$

where $D\psi_r$ and $D^2\psi_r$ are Euclidean gradient and Hessian of ψ_r. Moreover,

$$\psi_r(\boldsymbol{\theta}) = \psi_r(0) + r\big(D\psi(\mathbf{0}),\mathbf{L}(\boldsymbol{\theta})\big)_{\mathbb{R}^d} + \frac{r^2}{2}\mathrm{Trace}\big(D^2\psi(\mathbf{0})Q(\boldsymbol{\theta})\big) + r^3\tilde{\psi}_r(\boldsymbol{\theta})$$

where[5]

$$\begin{aligned}
\tilde{\psi}_r(\boldsymbol{\theta}) \equiv r^{-3}\int_{\mathbf{S}^{d-1}} P(\boldsymbol{\theta},\boldsymbol{\eta})\Big(&\psi(r\boldsymbol{\eta}) - \psi(0) - r\big(D\psi(\mathbf{0}),\boldsymbol{\eta}\big)_{\mathbb{R}^d} \\
&- \frac{r^2}{2}\big(D^2\psi(\mathbf{0})\boldsymbol{\eta},\boldsymbol{\eta}\big)_{\mathbb{R}^d}\Big)\,\bar{\lambda}_{\mathbf{S}^{d-1}}(d\boldsymbol{\eta}),
\end{aligned}$$

$$\mathbf{L}(\boldsymbol{\theta}) \equiv \int_{\mathbf{S}^{d-1}} \boldsymbol{\eta}P(\boldsymbol{\theta},\boldsymbol{\eta})\,\bar{\lambda}_{\mathbf{S}^{d-1}}(d\boldsymbol{\eta}) = \boldsymbol{\theta},$$

[5] The evaluation of the integrals below is done by solving the Dirichlet problem, not by direct computation.

and $Q = ((Q^{i,j}))$ with

$$Q^{i,j}(\boldsymbol{\theta}) \equiv \int_{\mathbf{S}^{d-1}} \eta^i \eta^j P(\boldsymbol{\theta}, \boldsymbol{\eta})\, \bar{\lambda}_{\mathbf{S}^{d-1}}(d\boldsymbol{\eta}) = \begin{cases} \theta^i \theta^j & \text{if } i \neq j \\ (\theta^i)^2 + d^{-1}\big(1 - |\boldsymbol{\theta}|^2_{\mathbb{R}^d}\big) & \text{if } i = j. \end{cases}$$

Because $\bar{\psi}_r$ and each of its derivatives is bounded independent of $r \in (0,1]$, we can now apply Corollary 9.29 to get

$$\begin{aligned} \frac{1}{2r^2}\mathbb{E}^{\mathbb{P}_x^M}&\left[\int_0^{\zeta_r(p)} \mathrm{Trace}\big((a - I_{\mathbb{R}^d})D^2\psi_r\big)\Big(r^{-1}\Theta_{\mathfrak{f}}\big(p(t)\big)\Big)\,dt\right] \\ &= -\frac{r^4}{6}\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\int_0^{\tau_1(\mathbf{w})} \mathrm{Trace}\Big(D^2\psi(\mathbf{0})\tilde{\Omega}\big(\mathbf{w}(t)\big)\,dt\right] + o(r^4), \end{aligned}$$

where

$$\tilde{\Omega}_{\mathfrak{f}}(\boldsymbol{\theta})^{i,j} \equiv \big(\Omega_{\mathfrak{f}}(\boldsymbol{\theta}, \mathbf{e}_i)\boldsymbol{\theta}, \mathbf{e}_j\big)_{\mathbb{R}^d} - d^{-1}\big(\mathfrak{R}_{\mathfrak{f}}\boldsymbol{\theta}, \boldsymbol{\theta}\big)_{\mathbb{R}^d}.$$

At the same time (again by Corollary 9.29)

$$\begin{aligned} \frac{1}{2r}\mathbb{E}^{\mathbb{P}_x^M}&\left[\int_0^{\zeta_r(p)} \big(\mathbf{b}, D\psi_r\big)_{\mathbb{R}^d}\Big(\big(r^{-1}\Theta_{\mathfrak{f}}\big(p(t)\big)\Big)\,dt\right] \\ &= -\frac{r^3}{3}\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\int_0^{\tau_1(\mathbf{w})} \Big(\mathfrak{R}_{\mathfrak{f}}\mathbf{w}(t), D\psi(\mathbf{0})\Big)_{\mathbb{R}^d}\,dt\right] + o(r^4). \end{aligned}$$

Hence, since (cf. (9.33))

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\int_0^{\tau_1(\mathbf{w})} \mathbf{w}(t)\,dt\right] = \mathbf{0} \text{ and } \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\int_0^{\tau_1(\mathbf{w})} \mathbf{w}(t)\otimes\mathbf{w}(t)\,dt\right] = \frac{I_{\mathbb{R}^d}}{2d(d+2)}$$

while, by (9.5), $D^2\psi(\mathbf{0}) = \mathrm{Hess}_{\mathfrak{f}}^M f$, we conclude that

$$\begin{aligned} &\mathbb{E}^{\mathbb{P}_x^M}\big[f\big(p(\zeta_r)\big)\big] - \int_{\mathbf{S}^{d-1}} f\circ\exp_{\mathfrak{f}}(\boldsymbol{\eta})\,\bar{\lambda}_{\mathbf{S}^{d-1}}(d\boldsymbol{\eta}) \\ &\qquad = \frac{r^4}{12d(d+2)}\mathrm{Trace}\big(\mathfrak{R}_{\mathfrak{f}}^{\#}\mathrm{Hess}_{\mathfrak{f}}^M f\big) + o(r^4), \end{aligned} \tag{9.34}$$

where

$$\mathfrak{R}_{\mathfrak{f}}^{\#} \equiv \mathfrak{R}_{\mathfrak{f}} - \frac{\mathrm{Trace}(\mathfrak{R}_{\mathfrak{f}})}{d} I_{\mathbb{R}^d}$$

is the *traceless Ricci curvature.*

Notice that (9.32) and (9.34) together say that the entire Ricci curvature is the encoded in the way that Brownian motion leaves small balls. The articles, alluded to earlier, by Pinsky and others examine what other geometric information can be recovered by looking at Brownian paths through a microscope.

CHAPTER 10

Perturbing Brownian Paths

The purpose of this chapter to develop a ubiquitous procedure for perturbing Brownian paths in order to achieve various goals. In particular, the procedure will allow us to give a derivation of Bismut's and related formulae without recourse to the information contained in § 6.2.1. Of course, the contents of § 6.1 were also based on a perturbation procedure: the Jacobian process $J(\cdot, x, \mathbf{w})$ there resulted from perturbing $p(\cdot, x, \mathbf{w})$ with respect to the initial point x. However, as we saw in Lemma 6.11, control over the size of this perturbation required us to control the largest eigenvalue of the highly non-intrinsic (cf. the closing Remark in § 6.1.2) transformation C_x in (6.12). The virtue of the procedure which we will develop here is that we will have much greater, in fact (cf. § 10.3.2 below) deterministic, control over the perturbation, and the control will be in terms of entirely intrinsic quantities.

Throughout, M will be a connected, complete, separable, d-dimensional Riemannian manifold, and $\mathcal{O}(M)$ will be its associated bundle of orthonormal frames.

10.1 Heuristic Explanation

Because it provides whatever geometric insight there is to be found, we will start by describing our perturbation procedure when the paths are smooth. Thus, throughout this discussion, $\mathbf{w}$ *will* be a "smooth $\mathbb{R}^d$-*valu*ed Brownian path" (i.e., a smooth element of $\mathfrak{W}(\mathbb{R}^d)$) and, for each $\mathfrak{f} \in \mathcal{O}(M)$, $\mathfrak{p}(\cdot, \mathfrak{f}, \mathbf{w})$ will be the horizontal lift of the path obtained by rolling $\mathbf{w}$ through $\mathfrak{f}$. That is, $\mathfrak{p}(\cdot, \mathfrak{f}, \mathbf{w})$ is determined by (8.27).

Now suppose that we were to imbed $\mathfrak{p}(\cdot, \mathfrak{f}, \mathbf{w})$ in a family $s \in [0,1] \longmapsto \mathfrak{P}_s(\cdot, \mathfrak{f}, \mathbf{w}) \in \mathcal{P}\big(\mathcal{O}(M)\big)$ in such a way that

$$\mathfrak{P}_0(\cdot, \mathfrak{f}, \mathbf{w}) = \mathfrak{p}(\cdot, \mathfrak{f}, \mathbf{w}) \text{ and } \mathfrak{P}_s(\cdot, \mathfrak{f}, \mathbf{w}) \text{ is horizontal for all } s \in [0,1].$$

Equivalently, we are assuming that there exists a maps $s \in [0,1] \longmapsto \mathfrak{f}_s \in \mathcal{O}(M)$ and $s \in [0,1] \longmapsto \mathbf{W}_s(\cdot, \mathfrak{f}, \mathbf{w}) \in \mathfrak{W}(\mathbb{R}^d)$ such that $\mathfrak{f}_0 = \mathfrak{f}$, $\mathbf{W}_0(\cdot, \mathfrak{f}, \mathbf{w}) = \mathbf{w}$, and

$$\mathfrak{P}_s(t, \mathfrak{f}, \mathbf{w}) = \mathfrak{p}\big(t, \mathfrak{f}_s, \mathbf{W}_s(\cdot, \mathfrak{f}, \mathbf{w})\big), \quad t \in [0, \infty).$$

Of course, $\mathfrak{f}_s = \mathfrak{P}_s(0, \mathfrak{f}, \mathbf{w})$ and (cf. (8.16)) $\dot{\mathbf{W}}_s(t, \mathfrak{f}, \mathbf{w}) = \phi\big(\dot{\mathfrak{P}}_s(t, \mathfrak{f}, \mathbf{w})\big)$. Thus, if we set[1] (cf. (8.16) and (8.17))

$$\Xi_s(t,\mathfrak{f},\mathbf{w}) = \phi\big(\mathfrak{P}'_s(t,\mathfrak{f},\mathbf{w})\big) \quad\text{and}\quad A_s(t,\mathfrak{f},\mathbf{w}) = \omega\big(\mathfrak{P}'_s(t,\mathfrak{f},\mathbf{w})\big),$$

then, because $\omega\big(\dot{\mathfrak{P}}_s(t,\mathfrak{f},\mathbf{w})\big) \equiv 0$, Cartan's structural equations (cf. especially (8.48) and (8.49)) say that

$$\begin{aligned}\frac{d\dot{\mathbf{W}}_s(t,\mathfrak{f},\mathbf{w})}{ds} &= \dot{\Xi}_s(t,\mathfrak{f},\mathbf{w}) - A_s(t,\mathfrak{f},\mathbf{w})\dot{\mathbf{W}}_s(t,\mathfrak{f},\mathbf{w})\\ \dot{A}_s(t,\mathfrak{f},\mathbf{w}) &= \Omega_{\mathfrak{P}_s(t,\mathfrak{f},\mathbf{w})}\big(\dot{\mathbf{W}}_s(t,\mathfrak{f},\mathbf{w}), \Xi_s(t,\mathfrak{f},\mathbf{w})\big).\end{aligned} \tag{10.1}$$

So far, we have not said what it is that we want our perturbation to achieve, and in this heuristic introduction we will avoid getting too specific. Nonetheless, we will insist that all our perturbations have in common the property that existence of and uniform control over the size of $\dot{\Xi}_s(t,\mathfrak{f},\mathbf{w})$ do not depend on the smoothness[2] of $\mathbf{w}$. The discussion which follows is an attempt to explain the price which we have to pay for this property. In particular, we will want to find out how far from that of $\mathbf{w}\rightsquigarrow \mathfrak{p}(\,\cdot\,,\mathfrak{f}_s,\mathbf{w})$ the $\mu_{\mathbb{R}^d}$-distribution $\pi\circ\mathfrak{P}_s(\,\cdot\,,\mathfrak{f},\mathbf{w})$ will have be. Equivalently, we will try to understand the distribution of $\mathbf{w}\rightsquigarrow\mathbf{W}_s(\,\cdot\,,\mathfrak{f},\mathbf{w})$ under $\mu_{\mathbb{R}^d}$. More precisely, given a $\varphi\in C^\infty_{\mathrm{c}}(\mathbb{R}^d;\mathbb{R})$, we are going to see what has to be subtracted from $\varphi\big(\mathbf{W}_s(t,\mathfrak{f},\mathbf{w})\big)$ in order to make the difference into a $\mu_{\mathbb{R}^d}$-martingale. Alternatively, we want to "compute"

$$\lim_{h\searrow 0}\frac{1}{h}\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\varphi\big(\mathbf{W}_s(t+h,\mathfrak{f},\mathbf{w})\big) - \varphi\big(\mathbf{W}_s(t,\mathfrak{f},\mathbf{w})\big)\,\Big|\,\overline{\mathcal{B}}_t\Big].$$

We begin by trying to "solve" (10.1) for $\mathbf{W}_s(\,\cdot\,,\mathfrak{f},\mathbf{w})$. To this end, determine $s\in[0,1]\longmapsto O_s(t,\mathfrak{f},\mathbf{w})\in O(\mathbb{R}^d)$, for each $t\ge 0$, by the equation

$$O'_s(t,\mathfrak{f},\mathbf{w}) + A_s(t,\mathfrak{f},\mathbf{w})O_s(t,\mathfrak{f},\mathbf{w}) = 0 \quad\text{with } O_0(t,\mathfrak{f},\mathbf{w}) = I.$$

Then, the first equation in (10.1), together with the initial condition $\mathbf{W}_0(\,\cdot\,,\mathfrak{f},\mathbf{w}) = \mathbf{w}$, is equivalent to

$$\dot{\mathbf{W}}_s(t,\mathfrak{f},\mathbf{w}) = O_s(t,\mathfrak{f},\mathbf{w})\dot{\mathbf{w}}(t) + \int_0^s O_{\sigma,s}(t,\mathfrak{f},\mathbf{w})\dot{\Xi}_\sigma(t,\mathfrak{f},\mathbf{w})\,d\sigma, \tag{*}$$

where $O_{\sigma,s}(t,\mathfrak{f},\mathbf{w}) \equiv O_s(t,\mathfrak{f},\mathbf{w})O_\sigma(t,\mathfrak{f},\mathbf{w})^\top$. Proceeding somewhat formally, but keeping in mind that, under $\mu_{\mathbb{R}^d}$, a typical path has centered, independent Gaussian increments which, over an interval of length dt, have covariance dtI (and, therefore, size $\sqrt{dt}$), we find that, up to $o(h)$-terms,

$$\begin{aligned}&\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\varphi\big(\mathbf{W}_s(t+h,\mathfrak{f},\mathbf{w})\big) - \varphi\big(\mathbf{W}_s(t,\mathfrak{f},\mathbf{w})\big)\,\Big|\,\overline{\mathcal{B}}_t\Big]\\ &= \mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\Big(\nabla\varphi\big(\mathbf{W}_s(t,\mathfrak{f},\mathbf{w})\big), \Delta_{[t,t+h]}\mathbf{W}_s(\,\cdot\,,\mathfrak{f},\mathbf{w})\Big)_{\mathbb{R}^d}\,\Big|\,\overline{\mathcal{B}}_t\Big]\\ &\quad+\tfrac12\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\Big(\Delta_{[t,t+h]}\mathbf{W}_s(\,\cdot\,,\mathfrak{f},\mathbf{w}), \nabla^2\varphi\big(\mathbf{W}_s(t,\mathfrak{f},\mathbf{w})\big)\Delta_{[t,t+h]}\mathbf{W}_s(\,\cdot\,,\mathfrak{f},\mathbf{w})\Big)_{\mathbb{R}^d}\,\Big|\,\overline{\mathcal{B}}_t\Big]\end{aligned}$$

---

[1] We will use "prime" to denote derivatives with respect to the perturbation parameter s and will reserve "dot" for derivatives with respect to the time parameter t.

[2] Keep in mind that although we are, at the moment, pretending that Brownian paths are smooth, we are eventually going to have to deal with $\mu_{\mathbb{R}^d}$-typical Brownian paths.

where $\Delta_{[t,t+h]}\mathbf{W}_s(\,\cdot\,,\mathfrak{f},\mathbf{w}) \equiv \big(\mathbf{W}_s(t+h,\mathfrak{f},\mathbf{w}) - \mathbf{W}_s(t,\mathfrak{f},\mathbf{w})\big)$, and we use $\nabla\varphi$ and $\nabla^2\varphi$ to denote the (standard, i.e., Euclidean) gradient and Hessian of φ. Next, note that

$$\begin{aligned}\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\Big(\Delta_{[t,t+h]}\mathbf{W}_s(\,\cdot\,,\mathfrak{f},\mathbf{w}),\nabla^2\varphi\big(\mathbf{W}_s(t,\mathfrak{f},\mathbf{w})\big)\Delta_{[t,t+h]}\mathbf{W}_s(\,\cdot\,,\mathfrak{f},\mathbf{w})\Big)_{\mathbb{R}^d}\,\Big|\,\overline{\mathcal{B}}_t\Big]\\ = h\big[\Delta_{\mathbb{R}^d}\varphi\big]\big(\mathbf{W}_s(t,\mathfrak{f},\mathbf{w})\big) + o(h)\end{aligned}$$

while, by (*),

$$\begin{aligned}&\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\Big(\nabla\varphi\big(\mathbf{W}_s(t,\mathfrak{f},\mathbf{w})\big),\Delta_{[t,t+h]}\mathbf{W}_s(\,\cdot\,,\mathfrak{f},\mathbf{w})\Big)_{\mathbb{R}^d}\,\Big|\,\overline{\mathcal{B}}_t\Big]\\ &= \int_t^{t+h}\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\Big(\nabla\varphi\big(\mathbf{W}_s(t,\mathfrak{f},\mathbf{w})\big),O_s(\tau,\mathfrak{f},\mathbf{w})\dot{\mathbf{w}}(\tau)\Big)_{\mathbb{R}^d}\,\Big|\,\overline{\mathcal{B}}_t\Big]\,d\tau\\ &\qquad + h\left(\nabla\varphi\big(\mathbf{W}_s(t,\mathfrak{f},\mathbf{w})\big),\int_0^s O_{\sigma,s}(t,\mathfrak{f},\mathbf{w})\dot{\Xi}_\sigma(t,\mathfrak{f},\mathbf{w})\,d\sigma\right)_{\mathbb{R}^d} + o(h),\end{aligned}$$

where, in order to get the $o(h)$-estimate, we have made our first application of our smoothness assumption about $\Xi_s(\,\cdot\,,\mathfrak{f},\mathbf{w})$.

To go further, remember that $\mathbf{w}(t+h)-\mathbf{w}(t)$ has mean $\mathbf{0}$ and is independent of $\overline{\mathcal{B}}_t$, and use this to justify

$$\begin{aligned}&\int_t^{t+h}\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[O_s(\tau,\mathfrak{f},\mathbf{w})\dot{\mathbf{w}}(\tau)\,\Big|\,\overline{\mathcal{B}}_t\Big]\,d\tau\\ &\qquad = \int_t^{t+h}\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\big(O_s(\tau,\mathfrak{f},\mathbf{w}) - O_s(t,\mathfrak{f},\mathbf{w})\big)\,\dot{\mathbf{w}}(\tau)\,\Big|\,\overline{\mathcal{B}}_t\Big]\,d\tau\\ &\qquad\qquad = \iint\limits_{t\le\tau\le\tau'\le t+h}\mathbb{E}^{\mu_{\mathbb{R}^d}}\big[\dot{O}_s(\tau',\mathfrak{f},\mathbf{w})\dot{\mathbf{w}}(\tau)\,\big|\,\overline{\mathcal{B}}_t\big]\,d\tau\,d\tau'.\end{aligned}$$

At the same time, because

$$\frac{d\dot{O}_s(\tau',\mathfrak{f},\mathbf{w})}{ds} + A_s(\tau',\mathfrak{f},\mathbf{w})\dot{O}_s(\tau',\mathfrak{f},\mathbf{w}) = -\dot{A}_s(\tau',\mathfrak{f},\mathbf{w})O_s(\tau',\mathfrak{f},\mathbf{w})$$

with $\dot{O}_0(\tau',\mathfrak{f},\mathbf{w}) = 0$, we see (cf. the second equation in (10.1) and the notation in (*)) that $-\dot{O}_s(\tau',\mathfrak{f},\mathbf{w})$ is equal to

$$\int_0^s O_{\sigma,s}(\tau',\mathfrak{f},\mathbf{w})\Omega_{\mathfrak{P}_\sigma(\tau',\mathfrak{f},\mathbf{w})}\big(\dot{\mathbf{W}}_\sigma(\tau',\mathfrak{f},\mathbf{w}),\Xi_\sigma(\tau',\mathfrak{f},\mathbf{w})\big)O_\sigma(\tau',\mathfrak{f},\mathbf{w})\,d\sigma.$$

But, because we are assuming that $\dot{\Xi}_\sigma(t,\mathfrak{f},\mathbf{w})$ is bounded, (*) shows that the error resulting from the replacement of the preceding integral by

$$\int_0^s O_{\sigma,s}(\tau',\mathfrak{f},\mathbf{w})\Omega_{\mathfrak{P}_\sigma(\tau',\mathfrak{f},\mathbf{w})}\big(O_\sigma(\tau',\mathfrak{f},\mathbf{w})\dot{\mathbf{w}}(\tau'),\Xi_\sigma(\tau',\mathfrak{f},\mathbf{w})\big)O_\sigma(\tau',\mathfrak{f},\mathbf{w})\,d\sigma$$

will remain bounded as long as $t \le \tau' \le t+h$. Hence, we have now shown that, for any orthonormal basis $(\mathbf{e}_1, \dots, \mathbf{e}_d)$ in $\mathbb{R}^d$,

$$
\begin{aligned}
&-\int_t^{t+h} \mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[O_s(\tau,\mathfrak{f},\mathbf{w})\dot{\mathbf{w}}(\tau)\,\Big|\,\overline{\mathcal{B}}_t\Big]d\tau \\
&= \sum_{k,k'=1}^d \int_0^s O_{\sigma,s}(t,\mathfrak{f},\mathbf{w})\Omega_{\mathfrak{P}_\sigma(t,\mathfrak{f},\mathbf{w})}\big(O_\sigma(t,\mathfrak{f},\mathbf{w})\mathbf{e}_{k'},\Xi_\sigma(t,\mathfrak{f},\mathbf{w})\big)O_\sigma(t,\mathfrak{f},\mathbf{w})\mathbf{e}_k\,d\sigma \\
&\qquad \times \mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\iint\limits_{t\le\tau\le\tau'\le t+h} \big(\dot{\mathbf{w}}(\tau'),\mathbf{e}_{k'}\big)_{\mathbb{R}^d}\big(\dot{\mathbf{w}}(\tau),\mathbf{e}_k\big)_{\mathbb{R}^d}\,d\tau d\tau'\,\middle|\,\overline{\mathcal{B}}_t\right] + o(h).
\end{aligned}
$$

Finally, we compute this last expectation value by taking the limit as $n \to \infty$ of (cf. (3.17)):

$$
\begin{aligned}
&\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\iint\limits_{t\le\tau\le\tau'\le t+h} \big(\dot{\mathbf{w}}_n(\tau'),\mathbf{e}_{k'}\big)_{\mathbb{R}^d}\big(\dot{\mathbf{w}}_n(\tau),\mathbf{e}_k\big)_{\mathbb{R}^d}\,d\tau d\tau'\,\middle|\,\overline{\mathcal{B}}_t\right] \\
&= \sum_{1\le m<m'<2^n h} \mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\big(\Delta_{m',n}\mathbf{w},\mathbf{e}_{k'}\big)_{\mathbb{R}^d}\big(\Delta_{m,n}\mathbf{w},\mathbf{e}_k\big)_{\mathbb{R}^d}\Big] \\
&\quad + \frac{1}{2}\sum_{1\le m<2^n h} \mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\big(\Delta_{m,n}\mathbf{w},\mathbf{e}_{k'}\big)_{\mathbb{R}^d}\big(\Delta_{m,n}\mathbf{w},\mathbf{e}_k\big)_{\mathbb{R}^d}\Big] + O(2^{-n}),
\end{aligned}
$$

where $\Delta_{m,n}\mathbf{w} \equiv \mathbf{w}(T_{m,n}) - \mathbf{w}(T_{m-1,n})$. Since, after applying orthogonality considerations, one finds that this limit equals $\frac{h\delta_{k,k'}}{2}$, we are now guessing that, as $h \searrow 0$, (cf. (8.58))

$$
\begin{aligned}
&-\int_t^{t+h} \mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[O_s(\tau,\mathfrak{f},\mathbf{w})\dot{\mathbf{w}}(\tau)\,\Big|\,\overline{\mathcal{B}}_t\Big]d\tau \\
&= \frac{h}{2}\sum_{k=1}^d \int_0^s O_{\sigma,s}(t,\mathfrak{f},\mathbf{w})\Omega_{\mathfrak{P}_\sigma(t,\mathfrak{f},\mathbf{w})}\big(O_\sigma(t,\mathfrak{f},\mathbf{w})\mathbf{e}_k,\Xi_\sigma(t,\mathfrak{f},\mathbf{w})\big)O_\sigma(t,\mathfrak{f},\mathbf{w})\mathbf{e}_k\,d\sigma \\
&= -\frac{h}{2}\int_0^s O_{\sigma,s}(t,\mathfrak{f},\mathbf{w})\mathfrak{R}_{\mathfrak{P}_\sigma(t,\mathfrak{f},\mathbf{w})}\Xi_\sigma(t,\mathfrak{f},\mathbf{w})\,d\sigma
\end{aligned}
$$

up to terms of order $o(h)$.

After combining the preceding with our earlier calculations, we finally arrive at

$$
\begin{aligned}
\lim_{h\searrow 0}\frac{\mathbb{E}^{\mu_{\mathbb{R}^d}}\big[\varphi\big(\mathbf{W}_s(t+h,\mathfrak{f},\mathbf{w})\big)-\varphi\big(\mathbf{W}_s(t,\mathfrak{f},\mathbf{w})\big)\,\big|\,\overline{\mathcal{B}}_t\big]}{h} \\
= \tfrac{1}{2}\Delta_{\mathbb{R}^d}\varphi\big(\mathbf{W}_s(t,\mathfrak{f},\mathbf{w})\big) + \big(\nabla\varphi\big(\mathbf{W}_s(t,\mathfrak{f},\mathbf{w})\big),\mathbf{b}_s(t,\mathfrak{f},\mathbf{w})\big)_{\mathbb{R}^d},
\end{aligned}
\tag{10.2}
$$

where

$$
\mathbf{b}_s(t,\mathfrak{f},\mathbf{w}) \equiv \int_0^s O_{\sigma,s}(t,\mathfrak{f},\mathbf{w})\big(\dot{\Xi}_\sigma(t,\mathfrak{f},\mathbf{w}) + \tfrac{1}{2}\mathfrak{R}_{\mathfrak{P}_\sigma(t,\mathfrak{f},\mathbf{w})}\Xi_\sigma(t,\mathfrak{f},\mathbf{w})\big)\,d\sigma.
$$

In particular, we conclude that $\mathbf{w}\rightsquigarrow\mathbf{W}_s$ should be a $\mu_{\mathbb{R}^d}$-Brownian motion when $s\rightsquigarrow\Xi_s(\,\cdot\,,\mathfrak{f},\mathbf{w})$ satisfies

$$\dot{\Xi}_s(t,\mathfrak{f},\mathbf{w}) + \tfrac{1}{2}\mathfrak{R}_{\mathfrak{P}_s(t,\mathfrak{f},\mathbf{w})}\Xi_s(t,\mathfrak{f},\mathbf{w}) = 0. \tag{10.3}$$

10.2 Formulation as a Flow

The conclusion to be drawn from the preceding (somewhat protracted) rumination is that, given $\mathfrak{f}\in\mathcal{O}(M)$ and $\mathbf{w}\in\mathfrak{W}(\mathbb{R}^d)$, we must learn how to solve for

$$s\in[0,1]\longmapsto\big(\mathbf{W}_s(\,\cdot\,,\mathfrak{f},\mathbf{w}),\mathfrak{P}_s(\,\cdot\,,\mathfrak{f},\mathbf{w})\big)\in\mathfrak{W}(\mathbb{R}^d)\times\mathcal{P}\big(\mathcal{O}(M)\big)$$

satisfying the system

$$\begin{aligned}&\mathbf{W}_0(\,\cdot\,,\mathfrak{f},\mathbf{w})=\mathbf{w}\quad\text{and}\quad\mathfrak{P}_0(0,\mathfrak{f},\mathbf{w})=\mathfrak{f}\\ &\dot{\mathfrak{P}}_s(t,\mathfrak{f},\mathbf{w})=\mathfrak{E}\big(\dot{\mathbf{W}}_s(t,\mathfrak{f},\mathbf{w})\big)_{\mathfrak{P}_s(t,\mathfrak{f},\mathbf{w})}\\ &\phi\big(\mathfrak{P}'_s(t,\mathfrak{f},\mathbf{w})\big)=\Xi_s(t,\mathfrak{f},\mathbf{w}),\end{aligned}\tag{10.4}$$

where $s\rightsquigarrow\Xi_s(t,\mathfrak{f},\mathbf{w})$ is smooth and, for each $s\in[0,1]$, $(t,\mathbf{w})\longmapsto\Xi_s(t,\mathfrak{f},\mathbf{w})$ is a $\{\overline{\mathcal{B}}_t:t\ge0\}$-progressively measurable function such that $\Xi_s(\,\cdot\,,\mathfrak{f},\mathbf{w})$ is smooth for all $\mathbf{w}\in\mathfrak{W}(\mathbb{R}^d)$. In fact, we must learn how to solve (10.4) so well that we can do so even when $\mathbf{w}$ is a $\mu_{\mathbb{R}^d}$-typical path. That is, we want to show that, when $\mathbf{w}_n$ is defined as in (3.17), the limit $\lim_{n\to\infty}\big(\mathbf{W}_s(t,\mathfrak{f},\mathbf{w}_n),\mathfrak{P}_s(t,\mathfrak{f},\mathbf{w}_n)\big)$ exists, in an appropriate sense, for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w}$. With this in mind, it is sensible to begin by reformulating (10.4) in a way which will make it more apparently amenable to the methods introduced in § 2.3. In particular, we are seeking a formulation of (10.4) in terms of flows generated by vector fields.

10.2.1. Initial Reformulation. The first step is provided by the following simple application of Cartan's structural equations.

10.5 Lemma. *Suppose that $(s,t)\in[0,1]\times[0,\infty)\longmapsto\mathfrak{P}_s(t)\in\mathcal{O}(M)$ is smooth in s and piecewise smooth in t, and determine $(s,t)\in[0,1]\times[0,\infty)\longmapsto\boldsymbol{\eta}_s(t)\in\mathbb{R}^d$ so that, for each t,*

$$\boldsymbol{\eta}'_s(t)+\omega\big(\mathfrak{P}'_s(t)\big)\boldsymbol{\eta}_s(t)=\frac{d}{dt}\phi\big(\mathfrak{P}'_s(t)\big)\quad\text{with }\boldsymbol{\eta}_0(t)=\phi\big(\dot{\mathfrak{P}}_0(t)\big).$$

Then $\mathfrak{P}_s(\,\cdot\,)$ is horizontal for each $s\in[0,1]$ if and only if $\mathfrak{P}_0(\,\cdot\,)$ is horizontal and

$$\frac{d}{dt}\omega\big(\mathfrak{P}'_s(t)\big)=\Omega_{\mathfrak{P}_s(t)}\Big(\boldsymbol{\eta}_s(t),\phi\big(\mathfrak{P}'_s(t)\big)\Big),\tag{10.6}$$

in which case $\dot{\mathfrak{P}}_s(t)=\mathfrak{E}\big(\boldsymbol{\eta}_s(t)\big)_{\mathfrak{P}_s(t)}$.

PROOF: First suppose that $\mathfrak{P}_s(\,\cdot\,)$ is horizontal for each s. Then, by (8.48),

$$\frac{d}{ds}\phi(\dot{\mathfrak{P}}_s(t)) = \frac{d}{dt}\phi(\mathfrak{P}'_s(t)) - \omega(\mathfrak{P}'_s(t))\phi(\dot{\mathfrak{P}}_s(t)),$$

and so, by uniqueness, $\phi(\dot{\mathfrak{P}}_s(t)) = \boldsymbol{\eta}_s(t)$. In particular, $\dot{\mathfrak{P}}_s(t) = \mathfrak{E}(\boldsymbol{\eta}_s(t))_{\mathfrak{P}_s(t)}$ and (8.49) implies that (10.6) holds.

Conversely, suppose that $\mathfrak{P}_0(\,\cdot\,)$ is horizontal and that (10.6) holds. Then, by (8.48) and (8.49), $s \in [0,1] \longmapsto \big(\phi(\dot{\mathfrak{P}}_s(t)), \omega(\dot{\mathfrak{P}}_s(t))\big) \in \mathbb{R}^d \times o(\mathbb{R}^d)$ is determined for each $t \geq 0$ by its value at $s = 0$ together with the system of equations

$$\begin{aligned}
\frac{d}{ds}\phi(\dot{\mathfrak{P}}_s(t)) &= \frac{d}{dt}\phi(\mathfrak{P}'_s(t)) - \omega(\mathfrak{P}'_s(t))\phi(\dot{\mathfrak{P}}_s(t)) + \omega(\dot{\mathfrak{P}}_s(t))\phi(\mathfrak{P}'_s(t))\\
\frac{d}{ds}\omega(\dot{\mathfrak{P}}_s(t)) &= -\Big[\omega(\mathfrak{P}'_s(t)), \omega(\dot{\mathfrak{P}}_s(t))\Big] + \Omega_{\mathfrak{P}_s(t)}\Big(\phi(\mathfrak{P}'_s(t)), \phi(\dot{\mathfrak{P}}_s(t)) - \boldsymbol{\eta}_s(t)\Big).
\end{aligned}$$

Hence, since $s \rightsquigarrow (\boldsymbol{\eta}_s(t), 0)$ has the same value at $s = 0$ and satisfies the same system of equations, we conclude that $\phi(\dot{\mathfrak{P}}_s(t)) = \boldsymbol{\eta}_s(t)$ and $\omega(\dot{\mathfrak{P}}_s(t)) = 0$ for all $s \in [0,1]$. $\square$

As a consequence of Lemma 10.5, we now see that, for any smooth map

$$s \in [0,1] \longmapsto \big(\mathbf{W}_s(\,\cdot\,, \mathfrak{f}, \mathbf{w}), \mathfrak{P}_s(\,\cdot\,, \mathfrak{f}, \mathbf{w})\big) \in \mathfrak{W}(\mathbb{R}^d) \times \mathcal{P}(\mathcal{O}(M))$$

satisfies (10.4) if and only if it satisfies the system

$$\begin{aligned}
\dot{\mathfrak{P}}_0(t, \mathfrak{f}, \mathbf{w}) &= \mathfrak{E}(\dot{\mathbf{w}}(t))_{\mathfrak{P}_0(t,\mathfrak{f},\mathbf{w})} \quad \text{with } \mathfrak{P}_0(0, \mathfrak{f}, \mathbf{w}) = \mathfrak{f}\\
\mathfrak{P}'_s(t, \mathfrak{f}, \mathbf{w}) &= \mathfrak{E}(\Xi_s(t, \mathfrak{f}, \mathbf{w}))_{\mathfrak{P}_s(t,\mathfrak{f},\mathbf{w})} + \lambda(A_s(t, \mathfrak{f}, \mathbf{w}))_{\mathfrak{P}_s(t,\mathfrak{f},\mathbf{w})}\\
\dot{A}_s(t, \mathfrak{f}, \mathbf{w}) &= \Omega_{\mathfrak{P}_s(t,\mathfrak{f},\mathbf{w})}(\dot{\mathbf{W}}_s(t, \mathfrak{f}, \mathbf{w}), \Xi_s(t, \mathfrak{f}, \mathbf{w}))\\
\frac{d}{ds}\dot{\mathbf{W}}_s(t, \mathfrak{f}, \mathbf{w}) + A_s(t, \mathfrak{f}, \mathbf{w})\dot{\mathbf{W}}_s(t, \mathfrak{f}, \mathbf{w}) &= \dot{\Xi}_s(t, \mathfrak{f}, \mathbf{w})\\
&\quad \text{with } \mathbf{W}_0(\,\cdot\,, \mathfrak{f}, \mathbf{w}) = \mathbf{w}.
\end{aligned} \tag{10.7}$$

10.2.2. Formulation as a System of O.D.E.'s on Pathspace. The system (10.7) provides the key to interpreting solutions to (10.4) as integral curves of vector fields. Namely, although (10.7) contains equations involving derivatives in both s and t, the equations containing s-derivatives are "gentle," even when $\mathbf{w}$ is non-differentiable. In particular, the solutions to the second and fourth equations in (10.7) can be thought of as smooth functions of $s \rightsquigarrow \Xi_s(t, \mathfrak{f}, \mathbf{w})$ and $s \rightsquigarrow A_s(t, \mathfrak{f}, \mathbf{w})$; and, when they are so considered, the first and third equations are seen to determine a (possibly time dependent) flow on an appropriate state space.

To develop this point of view, we must begin by determining what *an appropriate state space* is. But, because we want to remove from (10.4) the explicit appearance of s-derivatives, it should be clear that we will need a state space in which it is possible to hide these derivatives. That is, we should make our flow take place on a *pathspace*. With this in mind, let E be a separable Hilbert space, set $W^{(0)}(E) = L^2([0,1];E)$, and, for $m \ge 1$, define $W^{(m)}(E)$ to be the Sobolev space of continuous maps $s \in [0,1] \longmapsto \psi_s \in E$ having m square-integrable derivatives. Clearly, $W^{(m)}(E)$ is itself a separable Hilbert space when we give it the norm

$$\|\psi\|_{W^{(m)}(E)} = \left(\sum_{\ell=0}^{m-1} \left| \frac{d^\ell \psi_s}{ds^\ell} \bigg|_{s=0} \right|_E^2 + \int_0^1 \left| \frac{d^m \psi_s}{ds^m} \right|_E^2 ds \right)^{\frac{1}{2}}.$$

Finally, for $m \ge 1$, we use $W^{(m)}\big(\mathcal{O}(M)\big)$ to denote the space of absolutely continuous maps $s \in [0,1] \longmapsto \mathfrak{F}_s \in \mathcal{O}(M)$ with the property that[3]

$$s \in [0,1] \longmapsto \big(\phi(\mathfrak{F}'_s), \omega(\mathfrak{F}'_s)\big) \in W^{(m-1)}\big(\mathbb{R}^d \times o(\mathbb{R}^d)\big).$$

Given $\mathfrak{f} \in \mathcal{O}(M)$, $a \in W^{(m)}\big(o(\mathbb{R}^d)\big)$, and a $\mathbf{v} \in W^{(m)}(\mathbb{R}^d)$, define $s \in [0,1] \longmapsto O_s(a) \in O(\mathbb{R}^d)$, and $s \in [0,1] \longmapsto \mathfrak{F}_s(\mathfrak{f}, a, \mathbf{v}) \in \mathcal{O}(M)$ so that

$$O'_s(a) + a_s O_s(a) = 0 \quad \text{with } O_0(a) = I, \tag{10.8}$$

and

$$\mathfrak{F}'_s(\mathfrak{f}, a, \mathbf{v}) = \mathfrak{E}(\mathbf{v}_s)_{\mathfrak{F}_s(\mathfrak{f},a,\mathbf{v})} + \lambda\big(a_s\big)_{\mathfrak{F}_s(\mathfrak{f},a,\mathbf{v})} \quad \text{with } \mathfrak{F}_0(\mathfrak{f}, a, \mathbf{v}) = \mathfrak{f}. \tag{10.9}$$

It is easy, but somewhat tedious, to check the maps $a \in W^{(m)}\big(o(\mathbb{R}^d)\big) \longmapsto O(a) \in W^{(m)}\big(\mathrm{Hom}(\mathbb{R}^d;\mathbb{R}^d)\big)$ and

$$(\mathfrak{f}, a, \mathbf{v}) \in \mathcal{O}(M) \times W^{(m)}\big(o(\mathbb{R}^d) \times \mathbb{R}^d\big) \longmapsto \mathfrak{F}(\mathfrak{f}, a, \mathbf{v}) \in W^{(m)}\big(\mathcal{O}(M)\big)$$

are both smooth maps. For instance, if a, $b \in W^{(m)}\big(o(\mathbb{R}^d)\big)$ are given and $D_s \equiv \frac{d}{d\alpha} O_s(a + \alpha b)\big|_{\alpha=0}$, then $D'_s + a_s D_s + b_s O_s(a) = 0$ and so

$$\begin{aligned} (\partial_b)_a O_s \equiv \frac{d}{d\alpha} O_s(a + \alpha b)\Big|_{\alpha=0} &= -\int_0^s O_{\sigma,s}(a) b_\sigma O_\sigma(a)\, d\sigma \\ &\text{where } O_{\sigma,s}(a) \equiv O_s(a) O_\sigma(a)^\top. \end{aligned} \tag{10.10}$$

10.11 LEMMA. *Given* $\mathfrak{f} \in \mathcal{O}(M)$, *a piecewise smooth* $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$, *and a continuously differentiable* $t \in [0,\infty) \longmapsto \Xi(t, \mathfrak{f}, \mathbf{w}) \in W^{(m)}(\mathbb{R}^d)$,

$$\begin{aligned} t \in [0,\infty) \longmapsto & \big(A(t, \mathfrak{f}, \mathbf{w}), \mathbf{W}(t, \mathfrak{f}, \mathbf{w}), \mathfrak{P}(t, \mathfrak{f}, \mathbf{w})\big) \\ & \in W^{(m)}\big(o(\mathbb{R}^d)\big) \times W^{(m)}(\mathbb{R}^d) \times W^{(m)}\big(\mathcal{O}(M)\big) \end{aligned}$$

[3] We think of $o(\mathbb{R}^d)$ as a subspace of $\mathrm{Hom}(\mathbb{R}^d;\mathbb{R}^d)$, which we turn into a Hilbert space by identifying it with $\mathbb{R}^{d^2}$. That is, we put the Hilbert–Schmidt norm on $\mathrm{Hom}(\mathbb{R}^d;\mathbb{R}^d)$.

satisfies (10.7) if and only if (cf. (10.9))

$$\mathfrak{P}_s(t,\mathfrak{f},\mathbf{w}) = \mathfrak{F}_s\big(\mathfrak{p}(t,\mathfrak{f},\mathbf{w}), A(t,\mathfrak{f},\mathbf{w}), \Xi(t,\mathfrak{f},\mathbf{w})\big) \tag{10.12}$$

where (cf. the notation in (10.10))

$$\begin{gathered}\dot{\mathfrak{p}}(t,\mathfrak{f},\mathbf{w}) = \mathfrak{E}\big(\dot{\mathbf{w}}(t)\big)_{\mathfrak{p}(t,\mathfrak{f},\mathbf{w})} \quad \text{with } \mathfrak{p}(0,\mathfrak{f},\mathbf{w}) = \mathfrak{f} \\ \dot{A}_s(t,\mathfrak{f},\mathbf{w}) = \Omega_{\mathfrak{F}_s(\mathfrak{p}(t,\mathfrak{f},\mathbf{w}),A(t,\mathfrak{f},\mathbf{w}),\Xi(t,\mathfrak{f},\mathbf{w}))}\big(\dot{\mathbf{W}}_s(t,\mathfrak{f},\mathbf{w}), \Xi_s(t,\mathfrak{f},\mathbf{w})\big) \\ \dot{\mathbf{W}}_s(t,\mathfrak{f},\mathbf{w}) = O_s\big(A(t,\mathfrak{f},\mathbf{w})\big)\dot{\mathbf{w}}(t) + \int_0^s O_{\sigma,s}\big(A(t,\mathfrak{f},\mathbf{w})\big)\dot{\Xi}_\sigma(t,\mathfrak{f},\mathbf{w})\,d\sigma,\end{gathered} \tag{10.13}$$

in which case

$$\begin{aligned}\dot{\mathfrak{P}}_s(t,\mathfrak{f},\mathbf{w}) &= \mathfrak{E}\Big(O_s\big(A(t,\mathfrak{f},\mathbf{w})\big)\dot{\mathbf{w}}(t)\Big)_{\mathfrak{P}_s(t,\mathfrak{f},\mathbf{w})} \\ &\quad + \mathfrak{E}\left(\int_0^s O_{\sigma,s}\big(A(t,\mathfrak{f},\mathbf{w})\big)\dot{\Xi}_\sigma(t,\mathfrak{f},\mathbf{w})\,d\sigma\right)_{\mathfrak{P}_s(t,\mathfrak{f},\mathbf{w})}\end{aligned} \tag{10.14}$$

and

$$\mathfrak{P}'_s(t,\mathfrak{f},\mathbf{w}) = \mathfrak{E}\big(\Xi_s(t,\mathfrak{f},\mathbf{w})\big)_{\mathfrak{P}_s(t,\mathfrak{f},\mathbf{w})} + \lambda\big(A_s(t,\mathfrak{f},\mathbf{w})\big)_{\mathfrak{P}_s(t,\mathfrak{f},\mathbf{w})}. \tag{10.15}$$

PROOF: Obviously, (10.12) is equivalent to the second equation in (10.7). Thus, to prove the asserted equivalence, all that remains is the observation that the third equation in (10.13) is equivalent to the fourth equation in (10.7). Finally, to check (10.14), we use the equivalence of (10.7) and (10.4) to see that

$$\dot{\mathfrak{P}}_s(t,\mathfrak{f},\mathbf{w}) = \mathfrak{E}\big(\dot{\mathbf{W}}_s(t,\mathfrak{f},\mathbf{w})\big)_{\mathfrak{P}_s(t,\mathfrak{f},\mathbf{w})},$$

and, of course, (10.15) is just the second equation in (10.7). □

As Lemma 10.11 makes clear, all that we lack is a description of how $t \in [0,\infty) \longmapsto \Xi(t,\mathfrak{f},\mathbf{w}) \in W^{(m)}(\mathbb{R}^d)$ is to evolve; and, taking a hint from the considerations at the end of § 10.1, we should expect that it will be important for us to control $\dot{\Xi}_s(t,\mathfrak{f},\mathbf{w}) + \frac{1}{2}\mathfrak{R}_{\mathfrak{P}_s(t,\mathfrak{f},\mathbf{w})}\Xi_s(t,\mathfrak{f},\mathbf{w})$. Indeed, one reasonable choice would be to insist that (10.3) should hold, which is tantamount to taking (cf. (8.73)) $\Xi_s(t,\mathfrak{f},\mathbf{w}) = \mathfrak{J}_{\mathfrak{P}_s(\,\cdot\,,\mathfrak{f},\mathbf{w})}(t)\Xi_s(0,\mathfrak{f},\mathbf{w})$. In fact, for our purposes, it will suffice to take $\Xi_s(t,\mathfrak{f},\mathbf{w})$ to be of the form

$$\Xi_s(t,\mathfrak{f},\mathbf{w}) = \mathfrak{J}_{\mathfrak{P}_s(\,\cdot\,,\mathfrak{f},\mathbf{w})}(t)\boldsymbol{\theta}(t), \tag{10.16}$$

where $\boldsymbol{\theta} : [0,\infty) \longrightarrow \mathbb{R}^d$ is smooth. Notice that, in this case,

$$\dot{\Xi}_s(t,\mathfrak{f},\mathbf{w}) + \tfrac{1}{2}\mathfrak{R}_{\mathfrak{P}_s(t,\mathfrak{f},\mathbf{w})}\Xi_s(t,\mathfrak{f},\mathbf{w}) = \mathfrak{J}_{\mathfrak{P}_s(\,\cdot\,,\mathfrak{f},\mathbf{w})}(t)\dot{\boldsymbol{\theta}}(t), \tag{10.17}$$

and that (10.3) corresponds to constant $\boldsymbol{\theta}$.

10.2.3. The State Space and Vector Fields. We are now ready to describe the solution to (10.4) in terms of the flow generated by a vector field on an appropriate pathspace. Namely, let $m \geq 1$ be given, and set

$$\mathcal{H}^{(m)}(M) = \mathcal{O}(M) \times W^{(m)}\big(o(\mathbb{R}^d)\big) \times W^{(m)}(\mathbb{R}^d) \times W^{(m)}\big(\mathrm{Hom}(\mathbb{R}^d;\mathbb{R}^d)\big). \tag{10.18}$$

Next, choose an orthonormal basis $(\mathbf{e}_1, \dots, \mathbf{e}_d)$ in $\mathbb{R}^d$, and, for given $t \geq 0$ and $\boldsymbol{\theta} \in C^\infty([0,\infty);\mathbb{R}^d)$, define

$$\begin{aligned} \mathfrak{F}(t,\boldsymbol{\theta}) &: \mathcal{H}^{(1)}(M) \longrightarrow W^{(1)}\big(\mathcal{O}(M)\big) \\ \mathfrak{R}(t,\boldsymbol{\theta}) &: \mathcal{H}^{(1)}(M) \longrightarrow W^{(1)}\big(\mathrm{Hom}(\mathbb{R}^d;\mathbb{R}^d)\big) \\ \mathbf{H}(t,\boldsymbol{\theta}) &: \mathcal{H}^{(1)}(M) \longrightarrow W^{(1)}(\mathbb{R}^d) \\ \Omega_k(t,\boldsymbol{\theta}) &: \mathcal{H}^{(1)}(M) \longrightarrow W^{(1)}\big(o(\mathbb{R}^d)\big) \end{aligned} \tag{10.19}$$

so that (cf. (10.9), (10.8), and (10.10))

$$\begin{aligned} \mathfrak{F}(t,\boldsymbol{\theta})_s(\mathfrak{f},a,\mathbf{v},j) &= \mathfrak{F}_s\big(\mathfrak{f},a,j\boldsymbol{\theta}(t)\big) \\ \mathfrak{R}(t,\boldsymbol{\theta})_s(\mathfrak{f},a,\mathbf{v},j) &= \mathfrak{R}_{\mathfrak{F}(t,\boldsymbol{\theta})_s(\mathfrak{f},a,\mathbf{v},j)} \\ \mathbf{H}(t,\boldsymbol{\theta})_s(\mathfrak{f},a,\mathbf{v},j) &= \int_0^s O_{\sigma,s}(a)\Big(j_\sigma\dot{\boldsymbol{\theta}}(t) - \tfrac{1}{2}\mathfrak{R}(t,\boldsymbol{\theta})_\sigma(\mathfrak{f},a,\mathbf{v},j)j_\sigma\boldsymbol{\theta}(t)\Big)\,d\sigma \\ \Omega_0(t,\boldsymbol{\theta})_s(\mathfrak{f},a,\mathbf{v},j) &= \Omega_{\mathfrak{F}(t,\boldsymbol{\theta})_s(\mathfrak{f},a,\mathbf{v},j)}\Big(\mathbf{H}(t,\boldsymbol{\theta})_s(\mathfrak{f},a,\mathbf{v},j), j_s\boldsymbol{\theta}(t)\Big) \\ \Omega_k(t,\boldsymbol{\theta})_s(\mathfrak{f},a,\mathbf{v},j) &= \Omega_{\mathfrak{F}(t,\boldsymbol{\theta})_s(\mathfrak{f},a,\mathbf{v},j)}\big(O_s(a)\mathbf{e}_k, j_s\boldsymbol{\theta}(t)\big), \quad 1 \leq k \leq d, \end{aligned}$$

for $s \in [0,1]$ and $(\mathfrak{f},a,\mathbf{v},j) \in \mathcal{H}^{(1)}(M)$. All these functions are smooth and, for every $m \geq 1$, each one of them maps $\mathcal{H}^{(m)}(M)$ smoothly into the appropriate $W^{(m)}$-space. Referring to the preceding, define the vector fields $\{\widehat{\mathfrak{X}_0(t,\boldsymbol{\theta})}, \dots, \widehat{\mathfrak{X}_d(t,\boldsymbol{\theta})}\}$ on $\mathcal{H}^{(m)}(M)$ so that

$$\begin{aligned} \widehat{\mathfrak{X}_0(t,\boldsymbol{\theta})}_{(\mathfrak{f},a,\mathbf{v},j)} = &\big(\partial_{\Omega_0(t,\boldsymbol{\theta})(\mathfrak{f},a,\mathbf{v},j)}\big)_a + \big(\partial_{\mathbf{H}(t,\boldsymbol{\theta})(\mathfrak{f},a,\mathbf{v},j)}\big)_{\mathbf{v}} \\ &- \tfrac{1}{2}\big(\partial_{\mathfrak{R}(t,\boldsymbol{\theta})(\mathfrak{f},a,\mathbf{v},j))j}\big)_j \end{aligned} \tag{10.20}$$

and, for $1 \leq k \leq d$,

$$\widehat{\mathfrak{X}_k(t,\boldsymbol{\theta})}_{(\mathfrak{f},a,\mathbf{v},j)} = \mathfrak{E}\big(O(a)\mathbf{e}_k\big)_{\mathfrak{f}} + \big(\partial_{\Omega_k(t,\boldsymbol{\theta})(\mathfrak{f},a,\mathbf{v},j)}\big)_a + \big(\partial_{O(a)\mathbf{e}_k}\big)_{\mathbf{v}}. \tag{10.21}$$

10.22 LEMMA. *Assume that (8.72) holds. Let $\boldsymbol{\theta} : [0,\infty) \longrightarrow \mathbb{R}^d$ be a smooth function, and define the time-dependent vector fields $t \rightsquigarrow \widehat{\mathfrak{X}_k(t,\boldsymbol{\theta})}$, $0 \leq k \leq d$, accordingly, as in (10.20) and (10.21). Then, for each piecewise smooth $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$ and $\mathfrak{f} \in \mathcal{O}(M)$, there is a unique piecewise smooth path*

$$t \in [0,\infty) \longmapsto \big(\mathfrak{p}(t), A(t), \mathbf{W}(t), J(t)\big) \in \mathcal{H}^{(1)}(M)$$

which is the integral curve of the time-dependent vector field

$$t \rightsquigarrow \widehat{\mathfrak{X}_0(t,\boldsymbol{\theta})} + \sum_{k=1} \big(\dot{\mathbf{w}}(t), \mathbf{e}_k\big)_{\mathbb{R}^d} \widehat{\mathfrak{X}_k(t,\boldsymbol{\theta})}$$

starting at $(\mathfrak{f}, 0, \mathbf{0}, I)$. *In fact, for each* $m \geq 1$, *this path is piecewise smooth as a map into* $\mathcal{H}^{(m)}(M)$. *Next, set*

$$\tag{10.23} \begin{aligned} \Xi_s(t) &= J_s(t)\boldsymbol{\theta}(t) \\ \mathfrak{P}_s(t) = \big(\mathfrak{F}(t,\boldsymbol{\theta})\big)_s\big(\mathsf{p}(t), A(t), \mathbf{W}(t), J(t)\big) &= \mathfrak{F}_s\big(\mathsf{p}(t), A(t), \Xi(t)\big). \end{aligned}$$

Then

$$\tag{10.24} \begin{aligned} \dot{\mathfrak{P}}_s(t) &= \mathfrak{E}\big(\dot{\mathbf{W}}_s(t)\big)_{\mathfrak{P}_s(t)} \\ \mathfrak{P}_s'(t) &= \mathfrak{E}\big(\Xi_s(t)\big)_{\mathfrak{P}_s(t)} + \lambda\big(A_s(t)\big)_{\mathfrak{P}_s(t)} \\ \frac{d}{ds}\dot{\mathbf{W}}_s(t) &= \dot{\Xi}_s(t) - A_s(t)\dot{\mathbf{W}}_s(t) \quad \textit{with } \mathbf{W}_0 = \mathbf{w} \\ \dot{A}_s(t) &= \Omega_{\mathfrak{P}_s(t)}\big(\dot{\mathbf{W}}_s(t), \Xi_s(t)\big) \\ \Xi_s(t) &= \mathfrak{J}_{\mathfrak{P}_s}(t)\boldsymbol{\theta}(t). \end{aligned}$$

In particular,

$$\tag{10.25} \begin{aligned} \big(\mathfrak{F}(t,\boldsymbol{\theta})_s\big)_* \widehat{\mathfrak{X}_0(t,\boldsymbol{\theta})}_{(\mathfrak{f},a,\mathrm{v},j)} &= \mathfrak{E}\Big(\big(\mathbf{H}(t,\boldsymbol{\theta})\big)_s(\mathfrak{f}, a, \mathrm{v}, j)\Big)_{\mathfrak{F}(t,\boldsymbol{\theta})_s(\mathfrak{f},a,\mathrm{v},j)} \\ \big(\mathfrak{F}(t,\boldsymbol{\theta})_s\big)_* \widehat{\mathfrak{X}_k(t,\boldsymbol{\theta})}_{(\mathfrak{f},a,\mathrm{v},j)} &= \mathfrak{E}\big(O_s(a)\mathbf{e}_k\big)_{\mathfrak{F}(t,\boldsymbol{\theta})_s(\mathfrak{f},a,\mathrm{v},j)}, \quad 1 \leq k \leq d. \end{aligned}$$

PROOF: It follows from Lemma 10.11 that $t \rightsquigarrow \big(\mathsf{p}(t), A(t), \mathbf{W}(t), J(t)\big)$ is an integral curve of

$$t \rightsquigarrow \widehat{\mathfrak{X}_0(t,\boldsymbol{\theta})} + \sum_{k=1} \big(\dot{\mathbf{w}}(t), \mathbf{e}_k\big)_{\mathbb{R}^d} \widehat{\mathfrak{X}_k(t,\boldsymbol{\theta})}$$

if and only if

$$\dot{J}_s(t) + \frac{1}{2}\mathfrak{R}(t,\boldsymbol{\theta})_s\big(\mathsf{p}(t), A(t), \mathbf{W}(t), J(t)\big) = 0$$

and

$$(s,t) \rightsquigarrow \big(\mathfrak{P}_s(t,\mathfrak{f},\mathbf{w}), A_s(t,\mathfrak{f},\mathbf{w}), \mathbf{W}_s(t,\mathfrak{f},\mathbf{w})\big)$$

satisfies (10.7) with $\Xi_s(t,\mathfrak{f},\mathbf{w}) = J_s(t)\boldsymbol{\theta}(t)$ when

$$\begin{pmatrix} \mathfrak{P}_s(t,\mathfrak{f},\mathbf{w}) \\ A_s(t,\mathfrak{f},\mathbf{w}) \\ \mathbf{W}_s(t,\mathfrak{f},\mathbf{w}) \end{pmatrix} \equiv \begin{pmatrix} \big(\mathfrak{F}(t,\boldsymbol{\theta})\big)_s\big(\mathsf{p}(t), A(t), \mathbf{W}(t), J(t)\big) \\ A_s(t) \\ \mathbf{W}_s(t) \end{pmatrix}.$$

Hence, for such an integral curve, (10.24) is clear. Moreover, by the first line of (10.24), applied with $\mathbf{w} = \mathbf{0}$,

$$\big(\mathfrak{F}(t,\boldsymbol{\theta})_s\big)_* \widehat{\mathfrak{X}_0(t,\boldsymbol{\theta})}_{(\mathfrak{f},A,\mathbf{W},\Xi)} = \dot{\mathfrak{P}}_s(t,\mathfrak{f},\mathbf{0}) = \mathfrak{E}\big(\mathbf{H}(t,\boldsymbol{\theta})_s(\mathfrak{f},a,\mathrm{v},j)\big)_{\mathfrak{F}(t,\boldsymbol{\theta})_s(\mathfrak{f},a,\mathrm{v},j)},$$

which gives the first line of (10.25). Similarly, given the preceding, one can get the second line of (10.25) by applying the first line of (10.24) with $\mathbf{w}_k(t) = t\mathbf{e}_k$, $t \geq 0$.

Turning to the question of existence of integral curves, first suppose that M is compact and $m = 1$. In this case, the existence of the integral curves presents no problems since then the vector fields satisfy a uniform Lipschitz condition on $\mathcal{H}^{(1)}(M)$. When M is compact and $m > 1$, one proceeds as in § 2.4 and uses the observation that the mth order derivatives are governed by vector fields which are linear in the mth order derivatives, even though these vector fields may grow polynomially in the derivatives of order $1 \leq \ell < m$. Finally, when M is not compact, one uses the second and last lines of (10.24) together with (8.74) and the fact that M is complete in order to reduce to the case when M is compact. □

10.2.4. Perturbed Brownian Motion. Our reason for wanting to have the perturbation scheme in (10.4) formulated in terms of a flow was that we were preparing to invoke the reasoning introduced in § 2.3 and extended in Chapter 3 in order to handle $\mu_{\mathbb{R}^d}$-generic $\mathbf{w}$'s. With Lemma 10.22, we are now ready to do so.

10.26 THEOREM. *Assume that (8.72) holds, and let* $\mathfrak{f} \in \mathcal{O}(M)$ *and a smooth* $\boldsymbol{\theta} : [0,\infty) \longrightarrow \mathbb{R}^d$ *be given. For each* $n \in \mathbb{N}$, *use* $\mathbf{w}_n$ *to denote the polygonal approximation to* $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$ *described in (3.17) and*

$$t \in [0,\infty) \longmapsto \big(\mathsf{p}(t,\mathfrak{f},\mathbf{w}_n), A(t,\mathfrak{f},\mathbf{w}_n), \mathbf{W}(t,\mathfrak{f},\mathbf{w}_n), J(t,\mathfrak{f},\mathbf{w}_n)\big) \in \bigcap_{m\geq 1} \mathcal{H}^{(m)}(M)$$

to denote the integral curve (guaranteed by Lemma 10.22) of

$$t \rightsquigarrow \widehat{\mathfrak{X}_0(t,\boldsymbol{\theta})} + \sum_{k=1}^{d} (\dot{\mathbf{w}}_n(t), \mathbf{e}_k)_{\mathbb{R}^d} \widehat{\mathfrak{X}_k(t,\boldsymbol{\theta})} \quad \text{starting at } (\mathfrak{f}, 0, \mathbf{0}, I).$$

Then there exists a $\{\vec{\mathcal{B}}_t : t \geq 0\}$*-progressively measurable map*

$$\begin{aligned}(t,\mathbf{w}) &\in [0,\infty) \times \mathfrak{W}(\mathbb{R}^d) \\ &\longmapsto \big(\mathsf{p}(t,\mathfrak{f},\mathbf{w}), A(t,\mathfrak{f},\mathbf{w}), \mathbf{W}(t,\mathfrak{f},\mathbf{w}), J(t,\mathfrak{f},\mathbf{w})\big) \in \bigcap_{m\geq 1} \mathcal{H}^{(m)}(M)\end{aligned}$$

such that, for each $m \geq 1$,

$$\begin{aligned}&\big(\mathsf{p}(\,\cdot\,,\mathfrak{f},\mathbf{w}_n), A(\,\cdot\,,\mathfrak{f},\mathbf{w}_n), \mathbf{W}(\,\cdot\,,\mathfrak{f},\mathbf{w}_n), J(\,\cdot\,,\mathfrak{f},\mathbf{w}_n)\big) \\ &\quad\longrightarrow \big(\mathsf{p}(\,\cdot\,,\mathfrak{f},\mathbf{w}), A(\,\cdot\,,\mathfrak{f},\mathbf{w}), \mathbf{W}(\,\cdot\,,\mathfrak{f},\mathbf{w}), J(\,\cdot\,,\mathfrak{f},\mathbf{w})\big) \text{ in } \mathcal{H}^{(m)}(M)\end{aligned}$$

for $\mu_{\mathbb{R}^d}$-almost every $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$. Moreover, the $\mu_{\mathbb{R}^d}$-distribution $\widehat{\mathbb{P}^{\theta}_{\mathfrak{f}}}$ of

$$\begin{aligned}&\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)\\ &\quad\longmapsto \big(\mathfrak{p}(\,\cdot\,,\mathfrak{f},\mathbf{w}), A(\,\cdot\,,\mathfrak{f},\mathbf{w}), \mathbf{W}(\,\cdot\,,\mathfrak{f},\mathbf{w}), J(\,\cdot\,,\mathfrak{f},\mathbf{w})\big) \in \mathcal{P}\left(\bigcap_{m\ge 1}\mathcal{H}^{(m)}(M)\right)\end{aligned}$$

solves the martingale problem starting at $(\mathfrak{f},0,\mathbf{0},I)$ *for*

$$t\rightsquigarrow\widehat{\mathcal{L}^{\theta}_t} \equiv \widehat{\mathfrak{X}_0(t,\theta)} + \frac{1}{2}\sum_{k=1}^{d}\widehat{\mathfrak{X}_k(t,\theta)}^2 \tag{10.27}$$

PROOF: We start by assuming that M is compact and $m = 1$, in which case the vector fields have bounded first order derivatives. Thus, the only reason for hesitating to simply apply the results in Chapters 2 and 3 is that $\mathcal{H}^{(1)}(M)$ is infinite dimensional. However, because the infinite dimensional part $W^{(1)}\big(o(\mathbb{R}^d)\big) \times W^{(1)}(\mathbb{R}^d) \times W^{(1)}\big(\mathrm{Hom}(\mathbb{R}^d;\mathbb{R}^d)\big)$ of $\mathcal{H}^{(1)}(M)$ is a Hilbert space, all the L^2-computations made in those chapters go through without a hitch (cf. Theorems 5.1.22 in [35]). When M is compact and $m > 1$, one has to take advantage of the fact, alluded to in the proof of Lemma 10.22, about the structure of the vector fields governing the mth order derivatives. Namely, because they are uniformly Lipschitz with respect to the mth order derivatives and polynomial in the lower order derivatives, the form of Burkholder's inequality in Lemma 2.51 allows one to repeat the argument which we used in § 2.4 when we were estimating derivatives with respect to the starting point.

When M is not compact, one has to take advantage of the fact that, by Theorem 8.62, the martingale problem for $\frac{1}{2}\Delta_M$ starting at $x = \pi\mathfrak{f}$ is well-posed. To this end, first define $\mathfrak{P}_s(t,\mathfrak{f},\mathbf{w}_n)$ as in (10.23). Clearly, there is nothing to worry about as long as $(s,t,n)\rightsquigarrow\mathfrak{P}_s(t,\mathfrak{f},\mathbf{w}_n)$ stays in a compact, and, by construction (cf. the second and fourth lines of (10.24)) and the estimate in (8.74), this will be the case as long as $(t,n)\rightsquigarrow\pi\circ\mathfrak{p}(t,\mathfrak{f},\mathbf{w}_n)$ remains in a compact. But, by Theorem 8.62, for each $T > 0$ and $\epsilon > 0$, we can find a compact K in M such that

$$\mu_{\mathbb{R}^d}\Big(\big\{\mathfrak{p}(t,\mathfrak{f},\mathbf{w}_n) : n \in \mathbb{N} \text{ and } t \in [0,T]\big\} \subseteq K\Big) \ge 1-\epsilon.$$

Given this information, we can again apply exactly the same sort of localization procedure as we used in §3.2 to handle the analogous situation there. □

10.28 COROLLARY. *Again assume that (8.72) holds. Given a smooth θ : $[0,\infty) \longrightarrow \mathbb{R}^d$ and $\mathfrak{f} \in \mathcal{O}(M)$, define $\widehat{\mathbb{P}^{\theta}_{\mathfrak{f}}}$ accordingly, as in Theorem 10.26. If $\mathfrak{P}_s$ is defined as in (10.23), then, $\widehat{\mathbb{P}^{\theta}_{\mathfrak{f}}}$-almost surely,*

$$\mathfrak{P}_s'(t) = \mathfrak{E}\big(\mathfrak{J}_{\mathfrak{P}_s}(t)\theta(t)\big)_{\mathfrak{P}_s(t)} + \lambda\big(A_s(t)\big)_{\mathfrak{P}_s(t)}, \quad (s,t) \in [0,1]\times[0,\infty). \tag{10.29}$$

Moreover, for each $s \in [0,1]$ and $\varphi \in C^2_c(\mathbb{R}^d \times \mathcal{O}(M))$,

$$\varphi(\mathbf{W}_s(t), \mathfrak{P}_s(t)) - \int_0^t \Big(\partial_{\mathbf{b}_s^{\theta}(\tau,\mathfrak{P},A)} + \mathfrak{E}(\mathbf{b}_s^{\theta}(\tau,\mathfrak{P},A)) \\ + \frac{1}{2}\sum_{k=1}^d \Big(\partial_{\mathbf{e}_k} + \mathfrak{E}(\mathbf{e}_k)\Big)^2\Big)\varphi(\mathbf{W}_s(\tau), \mathfrak{P}_s(\tau))\, d\tau$$

is a $\widehat{\mathbb{P}^{\theta}_{\mathfrak{f}}}$-martingale when

$$\mathbf{b}_s^{\theta}(\tau,\mathfrak{P},A) \equiv \int_0^s O_{\sigma,s}(A(\tau))\mathfrak{I}_{\mathfrak{P}_\sigma}(\tau)\dot{\theta}(\tau)\, d\sigma. \tag{10.30}$$

PROOF: Clearly, (10.29) is what one gets by combining the second and last lines of (10.24). Moreover, to prove the asserted martingale property, all that we have to do is verify that (cf. (10.27))

$$\begin{aligned}[\widehat{\mathcal{L}^{\theta}_{\tau}}\varphi]\Big(\mathbf{W}_s, \mathfrak{F}(\tau,\theta)_s(\mathsf{p}(\tau), A(\tau), \mathbf{W}(\tau), J(\tau))\Big) \\ = \Big(\partial_{\mathbf{b}_s^{\theta}(\tau,\mathfrak{P},A)} + \mathfrak{E}(\mathbf{b}_s^{\theta}(\tau,\mathfrak{P},A)) \\ + \frac{1}{2}\sum_{k=1}^d \Big(\partial_{\mathbf{e}_k} + \mathfrak{E}(\mathbf{e}_k)\Big)^2\Big)\varphi(\mathbf{W}_s(\tau), \mathfrak{P}_s(\tau)).\end{aligned} \tag{10.31}$$

But, by (10.25), we know that, for $1 \le k \le d$,

$$\widehat{\mathfrak{X}_k(\tau,\theta)}_{\mathfrak{f},a,\mathbf{v},j}\varphi(\mathbf{v}, \mathfrak{F}(\tau,\theta)_s(\mathfrak{f},a,\mathbf{v},j)) = \Big(\partial_{O_s(a)\mathbf{e}_k} + \mathfrak{E}(O_s(a)\mathbf{e}_k)\Big)_{(\mathbf{v},\mathfrak{f})}\varphi,$$

and so

$$\begin{aligned}\widehat{\mathfrak{X}_k(\tau,\theta)}^2\varphi\Big(\mathbf{W}_s, \mathfrak{F}(\tau,\theta)_s(\mathsf{p}(\tau), A(\tau), \mathbf{W}(\tau), J(\tau))\Big) \\ = \Big(\partial_{\mathbf{c}^{\theta}_{s,k}(\tau,\mathfrak{P},A)} + \mathfrak{E}(\mathbf{c}^{\theta}_{s,k}(\tau,\mathfrak{P},A))\Big)_{\mathbf{W}_s(\tau),\mathfrak{P}_s(\tau)} \\ + \Big(\partial_{O(A_s(\tau))\mathbf{e}_k} + \mathfrak{E}\Big(O_s(A(\tau))\mathbf{e}_k\Big)\Big)^2_{(\mathbf{W}_s(\tau),\mathfrak{P}_s(\tau))}\varphi,\end{aligned}$$

where (cf. (10.10) and (10.19)) $\mathbf{c}^{\theta}_{s,k}(\tau,\mathfrak{P},A)$ is taken equal to

$$-\int_0^s O_{\sigma,s}(A(\tau))\Omega_k(\tau,\theta)_\sigma(\mathsf{p}(\tau), A(\tau), \mathbf{W}(\tau), J(\tau))O_\sigma(A(\tau))\mathbf{e}_k\, d\sigma.$$

Finally,

$$-\sum_{k=1}^d \Omega_k(\tau,\theta)_\sigma(\mathsf{p}(\tau), A(\tau), \mathbf{W}(\tau), J(\tau))O_\sigma(A(\tau))\mathbf{e}_k = \mathfrak{R}_{\mathfrak{P}_s(\tau)}J(\tau)\theta(\tau),$$

while

$$\sum_{k=1}^{d}\Big(\partial_{O(A_s(\tau))\mathbf{e}_k} + \mathfrak{E}\big(O_s(A(\tau))\mathbf{e}_k\big)\Big)^2 = \sum_{k=1}^{d}\Big(\partial_{\mathbf{e}_k} + \mathfrak{E}(\mathbf{e}_k)\Big)^2.$$

Hence, we have now proved that

$$\frac{1}{2}\sum_{k=1}^{d}\widehat{\mathfrak{X}_k(\tau,\boldsymbol{\theta})}^2\varphi\Big(\mathbf{W}_s,\mathfrak{F}(\tau,\boldsymbol{\theta})_s\big(\mathsf{p}(\tau),A(\tau),\mathbf{W}(\tau),J(\tau)\big)\Big)$$
$$= \Big(\partial_{\mathrm{d}_s^{\boldsymbol{\theta}}(\tau,\mathfrak{P},A)} + \mathfrak{E}\big(\mathrm{d}_s^{\boldsymbol{\theta}}(\tau,\mathfrak{P},A)\big) + \frac{1}{2}\sum_{k=1}^{d}\big(\partial_{\mathbf{e}_k}+\mathfrak{E}(\mathbf{e}_k)\big)^2\Big)_{(\mathbf{W}_s(\tau),\mathfrak{P}_s(\tau))}\varphi,$$

where

$$\mathrm{d}_s^{\boldsymbol{\theta}}(\tau,\mathfrak{P},A) \equiv \frac{1}{2}\int_0^s O_{\sigma,s}\big(A(\tau)\big)\mathfrak{R}_{\mathfrak{P}_\sigma(\tau)}J(\tau)\boldsymbol{\theta}(\tau)\,d\sigma.$$

Since, by (10.25),

$$\widehat{\mathfrak{X}_0(\tau,\boldsymbol{\theta})}_{(\mathfrak{f},a,\mathbf{v},j)}\varphi\big(\mathbf{v},\mathfrak{F}(t,\boldsymbol{\theta})_s\big) = \Big(\partial_{\mathbf{H}(\tau,\boldsymbol{\theta})_s(\mathfrak{f},a,\mathbf{v},j)} + \mathfrak{E}\big(\mathbf{H}(\tau,\boldsymbol{\theta})_s(\mathfrak{f},a,\mathbf{v},j)\big)\Big)_{(\mathbf{v},\mathfrak{f})}\varphi$$

and (cf. the last line of (10.24))

$$\mathbf{b}_s^{\boldsymbol{\theta}}(\tau,\mathfrak{P},A) = \mathrm{d}_s^{\boldsymbol{\theta}}(\tau,\mathfrak{P},A) + \mathbf{H}(\tau,\boldsymbol{\theta})_s\big(\mathfrak{P}(\tau),A(\tau),\mathbf{W}(\tau),J(\tau)\big)$$

$\widehat{\mathbb{P}_\mathfrak{f}^{\boldsymbol{\theta}}}$-almost surely, this completes the proof. □

10.3 Bochner via Perturbation of Brownian Paths

We are now ready to give a derivation of (8.76) by perturbing Brownian paths. Throughout we will be assuming that M is not only complete and connected but also that (8.72) holds for some $\gamma \in \mathbb{R}$. In particular, by Theorem 8.62, the martingale problem for Bochner's Laplacian $\frac{1}{2}\Delta_B$ is well-posed on $\mathcal{O}(M)$.

10.3.1. A Generalization of Bochner's Identity. Notice that, as we predicted at the end of § 9.1, something remarkable happens when we choose $\Xi_s(t) = \mathfrak{J}_{\mathfrak{P}_s}(t)\boldsymbol{\xi}$ for some fixed $\boldsymbol{\xi} \in \mathbb{R}^d$. Namely, because this corresponds to taking $\boldsymbol{\theta}(t) = \boldsymbol{\xi}$, $t \in [0,\infty)$, in Corollary 10.28, the corresponding $\widehat{\mathbb{P}_\mathfrak{f}^{\boldsymbol{\theta}}}$ will have the property that, for each $s \in [0,1]$, the $\widehat{\mathbb{P}_\mathfrak{f}^{\boldsymbol{\theta}}}$-distribution of (cf. (10.23))

$$(\mathsf{p},A,\mathbf{W},J) \in \mathcal{P}\big(\mathcal{H}^{(1)}(M)\big) \longmapsto (\mathbf{W}_s,\mathfrak{P}_s) \in \mathcal{P}\big(\mathbb{R}^d \times \mathcal{O}(M)\big)$$

solves the martingale problem for

$$\frac{1}{2}\sum_{k=1}^{d}\big(\partial_{\mathbf{e}_k}+\mathfrak{E}(\mathbf{e}_k)\big)^2$$

starting at $\left(\mathbf{0}, \mathfrak{q}(\mathfrak{f},\boldsymbol{\xi})_s\right)$, where $s \rightsquigarrow \mathfrak{q}(\mathfrak{f},\boldsymbol{\xi})_s$ is determined by

$$\mathfrak{q}(\mathfrak{f},\boldsymbol{\xi})'_s = \mathfrak{E}(\boldsymbol{\xi})_{\mathfrak{q}(\mathfrak{f},\boldsymbol{\xi})_s} \quad \text{with } \mathfrak{q}(\mathfrak{f},\boldsymbol{\xi})_0 = \mathfrak{f}. \tag{10.32}$$

But clearly, a second solution to the same martingale problem is the $\mu_{\mathbb{R}^d}$-distribution of (cf. Theorem 8.33)

$$\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto \Big(\mathbf{w}, \mathfrak{p}\big(\,\cdot\,, \mathfrak{q}(\mathfrak{f},\boldsymbol{\xi})_s, \mathbf{w}\big)\Big) \in \mathcal{P}\big(\mathbb{R}^d \times \mathcal{O}(M)\big).$$

Hence, by uniqueness, we have now proved the following powerful variation on Bochner's theme.

10.33 THEOREM. *Given $\mathfrak{f} \in \mathcal{O}(M)$ and $\boldsymbol{\xi} \in \mathbb{R}^d$, determine $s \rightsquigarrow \mathfrak{q}(\mathfrak{f},\boldsymbol{\xi})_s$ by (10.32) and let (cf. Theorem 10.26) $\widehat{\mathbb{P}^{\mathcal{O}(M)}_{\mathfrak{f},\boldsymbol{\xi}}} = \widehat{\mathbb{P}^{\boldsymbol{\theta}}_{\mathfrak{f}}}$ when $\boldsymbol{\theta}(t) = \boldsymbol{\xi}$ for all $t \geq 0$. Then (10.29) holds $\widehat{\mathbb{P}^{\mathcal{O}(M)}_{\mathfrak{f},\boldsymbol{\xi}}}$-almost surely, and, for each $s \in [0,1]$, the $\widehat{\mathbb{P}^{\mathcal{O}(M)}_{\mathfrak{f},\boldsymbol{\xi}}}$-distribution of (cf. (10.23)) $(\mathbf{W}_s, \mathfrak{P}_s)$ is the same as the $\mu_{\mathbb{R}^d}$-distribution of*

$$\mathbf{w} \rightsquigarrow \Big(\mathbf{w}, \mathfrak{p}\big(\,\cdot\,, \mathfrak{q}(\mathfrak{f},\boldsymbol{\xi})_s, \mathbf{w}\big)\Big)$$

In particular, the $\widehat{\mathbb{P}^{\mathcal{O}(M)}_{\mathfrak{f},\boldsymbol{\xi}}}$-distribution of $\mathfrak{P}_s$ is $\mathbb{P}^{\mathcal{O}(M)}_{\mathfrak{q}(\mathfrak{f},\boldsymbol{\xi})_s}$, and so $\pi \circ \mathfrak{P}_s$ under $\widehat{\mathbb{P}^{\mathcal{O}(M)}_{\mathfrak{f},\boldsymbol{\xi}}}$ has the same distribution, namely $\mathbb{P}^M_{\pi\circ q(\mathfrak{f},\boldsymbol{\xi})_s}$, as a Brownian motion on M starting at $\pi \circ \mathfrak{q}(\mathfrak{f},\boldsymbol{\xi})_s$.

As an essentially immediate consequence of the preceding, one gets the following generalization of (8.76). Notice that, as distinguished from the earlier derivation, the present one does not depend on the information in § 6.2.1. In particular, we avoid here the use of Li and Yau's estimate (6.27).

10.34 COROLLARY. *Given $\boldsymbol{\xi} \in \mathbb{R}^d$ and $\mathfrak{p} \in \mathcal{P}\big(\mathcal{O}(M)\big)$, define the tangent vector $X(\mathfrak{J}\boldsymbol{\xi})_{\mathfrak{p}}$ to $\mathcal{P}(M)$ at $\pi \circ \mathfrak{p}$ so that*

$$X(\mathfrak{J}\boldsymbol{\xi})_{\mathfrak{p}} = \frac{d}{ds}p_s\Big|_{s=0} \quad \text{where } s \in [0,1] \longmapsto p_s \in \mathcal{P}\big(\mathcal{O}(M)\big) \tag{10.35}$$
$$\text{is continuously differentiable with } p_0'(t) = \mathfrak{p}(t)\mathfrak{J}_{\mathfrak{p}}(t)\boldsymbol{\xi}.$$

Given $\Phi \in C^1_{\mathrm{b}}\big(\mathcal{P}(M);\mathbb{R}\big)$, set $\varphi(x) = \mathbb{E}^{\mathbb{P}^M_x}[\Phi]$, $x \in M$. If either (cf. (8.72)) $\gamma \geq 0$ or Φ is $\mathcal{F}_T$ measurable for some $T \in [0,\infty)$, then $\varphi \in C^1_{\mathrm{b}}(M;\mathbb{R})$ and

$$\mathfrak{E}(\boldsymbol{\xi})_{\mathfrak{f}}\varphi \circ \pi = \mathbb{E}^{\mathbb{P}^{\mathcal{O}(M)}_{\mathfrak{f}}}\big[X(\mathfrak{J}\boldsymbol{\xi})_{\mathfrak{p}}\Phi\big]. \tag{10.36}$$

In particular, when $\Phi(p) = f \circ p(T)$, (10.36) is the same statement as the one in (8.76).

10.3.2. An Application to Coupling. Expanding a little on the comment made just before Cororllary 10.34, notice that the virtue of the present perturbation scheme, as distinguished from the one used in § 6.1, is that it gives us *pathwise* control. To emphasize this point, we now show that this control leads to a remarkable coupling (cf. footnote (3) at the beginning of § 9.4) of Brownian motions starting at two different points in M. Because of its similarity to a coupling which W. Kendell [23] (see also [6]) produced, albeit by an entirely different[4] line of reasoning, we call this *Kendell's coupling.*

10.37 THEOREM. *Given a pair of points x and y in M, let $s \in [0,1] \longmapsto q_s \in M$ be a minimal geodesic running from x to y. Then there is a probability measure $\mathbb{K}_{x,y}$ on the space $\mathcal{P}\big(C^\infty([0,1];M)\big)$ of continuous paths*

$$t \in [0,\infty) \longmapsto P(t) \in C^\infty\big([0,1];M\big)$$

such that, for each $s \in [0,1]$, the $\mathbb{K}_{x,y}$-distribution of $P_s(\,\cdot\,)$ is $\mathbb{P}^M_{q_s}$, and, $\mathbb{K}_{x,y}$-almost surely,

$$\|P_s'(t)\| \le e^{-\gamma t}\mathrm{dist}^M(x,y), \quad (s,t) \in [0,1]\times[0,\infty). \tag{10.38}$$

In particular, $\mathbb{K}_{x,y}$-almost surely:

$$\begin{gathered}\mathrm{dist}^M\big(P_{s_2}(t),P_{s_1}(t)\big) \le e^{-\gamma t}(s_2-s_1)\mathrm{dist}^M(x,y)\\ \textit{for all } 0 \le s_1 < s_2 \le 1 \textit{ and } t \in [0,\infty).\end{gathered} \tag{10.39}$$

PROOF: Choose $\mathfrak{f} \in \pi^{-1}x$, and set $\boldsymbol{\xi} = \mathfrak{f}^{-1}q_0'$. Then (cf. (10.32)) $s \rightsquigarrow \mathfrak{q}(\mathfrak{f},\boldsymbol{\xi})_s$ is the horizontal lift of q to $\mathfrak{f}$. Now take $\mathbb{K}_{x,y}$ to be the (cf. Theorem (10.32)) $\widehat{\mathbb{P}^{\mathcal{O}(M)}_{\mathfrak{f},\boldsymbol{\xi}}}$-distribution of $\pi\circ\mathfrak{P}$. Since, for each $s \in [0,1]$, $\mathfrak{P}_s$ under $\widehat{\mathbb{P}^{\mathcal{O}(M)}_{\mathfrak{f},\boldsymbol{\xi}}}$ solves the martingale problem for $\frac{1}{2}\Delta_B$ starting at $\mathfrak{q}(\mathfrak{f},\boldsymbol{\xi})_s$, there is no doubt that $P_s(\,\cdot\,)$, under $\mathbb{K}_{x,y}$, is a Brownian motion on M starting at q_s. Moreover, for each $t \in [0,\infty)$, $\pi_*\mathfrak{P}_s'(t) = \mathfrak{P}_s(t)\mathfrak{J}_{\mathfrak{P}_s}(t)\boldsymbol{\xi}$, and so, by (8.74),

$$\big\|\pi_*\mathfrak{P}_s'(t)\big\| \le e^{-\gamma t}|\boldsymbol{\xi}|_{\mathbb{R}^d} \quad \widehat{\mathbb{P}^{\mathcal{O}(M)}_{\mathfrak{f},\boldsymbol{\xi}}}\text{-almost surely,}$$

which is equivalent to (10.38) when one takes into account the fact that $|\boldsymbol{\xi}|_{\mathbb{R}^d} = \mathrm{dist}^M(x,y)$. Finally, to get (10.39) from (10.38), one simply integrates:

$$\mathrm{dist}^M\big(P_{s_2}(t),P_{s_1}(t)\big) \le \int_{s_1}^{s_2}\|P_s'(t)\|\,ds \le e^{-\gamma t}(s_2-s_1)\mathrm{dist}^M(x,y)$$

$\mathbb{K}_{x,y}$-almost surely. □

[4] Kendell's arguement is based on much more geometrically intuitive ideas than ours. Unfortunately, by the time his ideas are converted into rigorous mathematics, their intuitive appeal loses some of its lustre.

At least two comments should be made about the preceding version of Kendell's coupling. First, when $M = \mathbb{R}^d$, it, like his, is precisely the naive coupling which one gets by taking $P_s(t) = (1-s)x + sy + \mathbf{w}(t)$. In particular, for each t, the connecting path $s \rightsquigarrow P_s(t)$ is a geodesic. However, unless M is flat, our $s \rightsquigarrow P_s(t)$ will not be a geodesic when $t > 0$, and therefore the square of the true distance between $P_0(t)$ and $P_1(t)$ will be strictly less than the energy $\int_0^1 \|P_s'(t)\|^2\, ds$, which means, in turn, that the inequality in (10.39) will be a strict one. To check this assertion, we use (8.23) to see that

$$\mathfrak{P}_s(t)^{-1}\frac{D}{ds}P_s'(t) = \mathfrak{J}'_{\mathfrak{P}_s}(t)\boldsymbol{\xi} + A_s(t)\mathfrak{J}_{\mathfrak{P}_s}(t)\boldsymbol{\xi} \quad \text{where } \mathfrak{J}_s'(t) \equiv \frac{d}{ds}\mathfrak{J}_{\mathfrak{P}_s}(t).$$

Thus, even if $\mathfrak{J}_s'(t) \equiv 0$, as it will be when the Ricci curvature is constant (i.e. M is an Einstein manifold), the acceleration $\frac{D}{ds}P_s'(t)$ will not vanish identically unless $A_s(t)$ does, which is possible only if M has no curvature. A second comment is that ours, like all other versions of Kendell's coupling, gives dramatic verification of the ergodic property proved in § 6.1.4 when $\gamma > 0$. Indeed, (10.38) makes it clear that Kendell's coupling produces Brownian paths issuing from different points but converging at an exponential rate with rate constant at least γ.

10.4 Bismut via Perturbation of Brownian Paths

In this section we will show how a slight modification of the preceding perturbation scheme allows one to derive Bismut's formula (cf. (8.80)). Although it is somewhat obscured by the formalism, the basic idea is very simple. Namely, let $\boldsymbol{\theta} \in C^\infty([0,\infty);\mathbb{R}^d)$, set $\boldsymbol{\xi} = \boldsymbol{\theta}(0)$, and determine

$$(s,\mathfrak{f}) \in [0,1] \times \mathcal{O}(M) \longmapsto \mathfrak{q}(\mathfrak{f},\boldsymbol{\xi})_s \in \mathcal{O}(M)$$

by (10.32). What we want to show is that, for each $s \in [0,1]$ and $T \in [0,\infty)$, $\mathbb{P}^M_{q(\mathfrak{f},\boldsymbol{\xi})_s} \restriction \mathcal{F}_T$ is equivalent (i.e., mutually absolutely continuous with respect) to $\widehat{\mathbb{P}^{\boldsymbol{\theta}}_{\mathfrak{f}}}$-distribution of $\mathfrak{P}_s \restriction [0,T]$. Once we know this, we will use $E_s(T)$ to denote the Radon–Nikokym derivative arising from the equivalence and will write

$$\begin{aligned}\mathfrak{f}\boldsymbol{\xi}\mathbf{P}_T^M f &= \frac{d}{ds}\mathbb{E}^{\mathbb{P}^M_{\mathfrak{q}(\mathfrak{f},\boldsymbol{\xi})_s}}\Big[f\circ\pi\big(\mathfrak{p}(T)\big)\Big]\Big|_{s=0} = \frac{d}{ds}\mathbb{E}^{\widehat{\mathbb{P}}^{\boldsymbol{\theta}}_{\mathfrak{f}}}\Big[E_s(T)f\circ\pi\big(\mathfrak{P}_s(T)\big)\Big]\Big|_{s=0}\\ &= \mathbb{E}^{\widehat{\mathbb{P}}^{\boldsymbol{\theta}}_{\mathfrak{f}}}\Big[E_0'(T)\,f\circ\pi\big(\mathfrak{p}(T)\big)\Big] + \mathbb{E}^{\mathbb{P}^M_{\mathfrak{f}}}\Big[\mathfrak{E}\big(\mathfrak{J}_{\mathfrak{p}}(T)\boldsymbol{\theta}(T)\big)_{\mathfrak{p}(T)}f\circ\pi\Big].\end{aligned}$$

In particular, by insisting that $\boldsymbol{\theta}(T) = \mathbf{0}$, we will get Bismut's formula with a Bismut factor given by $E_0'(T)$. In other words, we plan to derive Bismut's formula via an *integration by parts* technique, very much like the one described in § 1.2.6.

Even though we may not always mention it, we will again be assuming that (8.72) holds throughout.

10.4.1. The Perturbation and the Radon–Nikodym Factor. Suppose that we take $\boldsymbol{\theta} \in C^\infty([0,\infty);\mathbb{R}^d)$ to be non-constant. Then (even if M is flat) it will be no longer true that the distribution of $\mathfrak{P}_s$ under $\widehat{\mathbb{P}^{\boldsymbol{\theta}}_{\mathfrak{f}}}$ will be $\mathbb{P}^M_{\mathfrak{q}(\mathfrak{f},\mathfrak{E})_s}$. Indeed, according to Corollary 10.28, there is an extra first order term (cf. (10.30)) required to make $\varphi(\mathfrak{P}_s(t))$ into a $\widehat{\mathbb{P}^{\boldsymbol{\theta}}_{\mathfrak{f}}}$-martingale. The idea is to eliminate this extra term by the introduction of an appropriate factor, namely, the Radon–Nikodym factor alluded to above. In fact, our goal here is to show that the factor (cf. (8.77) and (10.23))

$$
\begin{aligned}
E^{\boldsymbol{\theta}}_s(t,\mathbf{W},\mathfrak{P}) &\equiv \exp\left(-\int_0^s \big[B(\boldsymbol{\theta})\big](t,\mathbf{W}_\sigma,\mathfrak{P}_\sigma)\,d\sigma\right)\\
\text{with } \big[B(\boldsymbol{\theta})\big](t,\mathbf{w},\mathfrak{p}) &\equiv \int_0^t \Big(\mathfrak{J}_{\mathfrak{p}}(\tau)\dot{\boldsymbol{\theta}}(\tau), d\mathbf{w}(\tau)\Big)_{\mathbb{R}^d}.
\end{aligned}
\tag{10.40}
$$

serves.

The key to everything which follows is our ability to get a handle on the $\widehat{\mathbb{P}^{\boldsymbol{\theta}}_{\mathfrak{f}}}$-distribution of

$$
(\mathfrak{p},A,\mathbf{W},J) \in \mathcal{P}\big(\mathcal{H}^{(1)}(M)\big) \longmapsto \big(\mathfrak{P}_s,\rho_s\big) \in \mathcal{P}\big(\mathcal{O}(M)\times\mathbb{R}\big),
$$

where $\mathfrak{P}$ is given by (10.23) and

$$
\rho_s(t) = \int_0^t \big(J_s(\tau)\dot{\boldsymbol{\theta}}(\tau), d\mathbf{W}_s(\tau)\big)_{\mathbb{R}^d}. \tag{10.41}
$$

For this purpose, it is best to go back to the approximation procedure discussed in Theorem 10.26. One then realizes that, because the integrals involved are defined in the sense of Riemann–Stieltjes, $\mu_{\mathbb{R}^d}$-almost surely one has that

$$
\begin{aligned}
\rho_s(t,\mathfrak{f},\mathbf{w}_n) &\equiv \int_0^t \big(J_s(\tau,\mathfrak{f},\mathbf{w}_n)\dot{\boldsymbol{\theta}}(\tau), \dot{\mathbf{W}}_s(\tau,\mathfrak{f},\mathbf{w}_n)\big)_{\mathbb{R}^d}\,d\tau\\
&\longrightarrow \rho_s(t,\mathfrak{f},\mathbf{w}) \equiv \int_0^t \big(J_s(\tau,\mathfrak{f},\mathbf{w})\dot{\boldsymbol{\theta}}(\tau), d\mathbf{W}_s(\tau,\mathfrak{f},\mathbf{w})\big)_{\mathbb{R}^d}
\end{aligned}
$$

uniformly as (s,t) runs over compacts. The importance of this observation stems from the easily checked fact that, for each $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$ and $n \in \mathbb{N}$,

$$
\begin{aligned}
t \in [0,\infty) \longmapsto &\big(\mathfrak{p}(t,\mathfrak{f},\mathbf{w}_n), A(t,\mathfrak{f},\mathbf{w}_n), \mathbf{W}(t,\mathfrak{f},\mathbf{w}_n), \rho(t,\mathfrak{f},\mathbf{w}_n)\big)\\
&\in \bigcap_{m\ge 1}\big(\mathcal{H}^{(m)}(M)\times W^{(m)}(\mathbb{R})\big)
\end{aligned}
$$

is the integral curve of the time-dependent vector field

$$
t \rightsquigarrow \widetilde{\mathfrak{X}_0(t,\boldsymbol{\theta})} + \sum_{k=1}^d \big(\dot{\mathbf{w}}_n(t),\mathbf{e}_k\big)_{\mathbb{R}^d}\widetilde{\mathfrak{X}_k(t,\boldsymbol{\theta})},
$$

where (cf. (10.19) and (10.20))

$$\widetilde{\mathfrak{X}_0(t,\boldsymbol{\theta})}_{(\mathfrak{f},a,\mathbf{v},j,\rho)} \equiv \widehat{\mathfrak{X}_0(t,\boldsymbol{\theta})}_{(\mathfrak{f},a,\mathbf{v},j)} + \left(\partial_{(j\dot{\boldsymbol{\theta}}(t),\mathbf{H}(t,\boldsymbol{\theta})(\mathfrak{f},a,\mathbf{v},j))_{\mathbb{R}^d}}\right)_\rho \tag{10.42}$$

and (cf. (10.21))

$$\widetilde{\mathfrak{X}_k(t,\boldsymbol{\theta})}_{(\mathfrak{f},a,\mathbf{v},j,\rho)} \equiv \widehat{\mathfrak{X}_k(t,\boldsymbol{\theta})}_{(\mathfrak{f},a,\mathbf{v},j)} + \left(\partial_{(j\dot{\boldsymbol{\theta}}(t),O(a)\mathbf{e}_k)_{\mathbb{R}^d}}\right)_\rho \tag{10.43}$$

for $1 \le k \le d$. Indeed, after combining this with the preceding, we conclude that the $\mu_{\mathbb{R}^d}$-distribution of

$$\begin{aligned} \mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto & \big(\mathfrak{p}(t,\mathfrak{f},\mathbf{w}), A(t,\mathfrak{f},\mathbf{w}), \mathbf{W}(t,\mathfrak{f},\mathbf{w}), \rho(t,\mathfrak{f},\mathbf{w})\big) \\ & \in \bigcap_{m\ge 1} \mathcal{P}\big(\mathcal{H}^{(m)}(M) \times W^{(m)}(\mathbb{R})\big) \end{aligned}$$

solves the martingale problem for the time-dependent operator (cf. (10.42) and (10.43))

$$t \rightsquigarrow \widetilde{\mathcal{L}_t^{\boldsymbol{\theta}}} \equiv \widetilde{\mathfrak{X}_0(t,\boldsymbol{\theta})} + \frac{1}{2}\sum_{k=1}^d \widetilde{\mathfrak{X}_k(t,\boldsymbol{\theta})}^2. \tag{10.44}$$

Hence, since $J_s(t,\mathfrak{f},\mathbf{w}) = \mathfrak{J}_{\mathfrak{P}_s(\,\cdot\,,\mathfrak{f},\mathbf{w})}(t)$ and therefore

$$\big[B(\boldsymbol{\theta})\big]\big(t, \mathbf{W}_s(\,\cdot\,,\mathfrak{f},\mathbf{w}), \mathfrak{P}_s(\,\cdot\,,\mathfrak{f},\mathbf{w})\big) = \rho_s(t,\mathfrak{f},\mathbf{w}) \quad \mu_{\mathbb{R}^d}\text{-almost surely,}$$

we have now verified the following statement.

10.45 THEOREM. *Assume that (8.72) holds, let $\boldsymbol{\theta} \in C^\infty\big([0,\infty);\mathbb{R}^d\big)$ and $\mathfrak{f} \in \mathcal{O}(M)$ be given, and define $\widehat{\mathbb{P}_{\mathfrak{f}}^{\boldsymbol{\theta}}}$ accordingly, as in Theorem 10.26. Then the $\widehat{\mathbb{P}_{\mathfrak{f}}^{\boldsymbol{\theta}}}$-distribution $\widetilde{\mathbb{P}_{\mathfrak{f}}^{\boldsymbol{\theta}}}$ of*

$$\begin{aligned} (\mathfrak{p}, A, \mathbf{W}, J) \in \bigcap_{m\ge 1} \mathcal{P}\big(\mathcal{O}(M)\big) \longmapsto & \big(\mathfrak{p}, A, \mathbf{W}, J, [B(\boldsymbol{\theta})](\,\cdot\,,\mathbf{W},\mathfrak{P})\big) \\ & \in \bigcap_{m\ge 1} \mathcal{P}\big(\mathcal{H}^{(m)}(M) \times W^{(m)}(\mathbb{R})\big) \end{aligned}$$

solves the martingale problem for the $t \rightsquigarrow \widetilde{\mathcal{L}_t^{\boldsymbol{\theta}}}$ in (10.44) starting at $(\mathfrak{f}, 0, \mathbf{0}, I, 0)$.

Our first application of Theorem 10.45 shows that the use of (cf. (10.40)) $E_s^{\boldsymbol{\theta}}(T)$ as a Radon–Nikodym factor leads to a consistently defined probability measure on $\mathcal{P}\big(\mathcal{O}(M)\big)$.

10.46 LEMMA. *Assume that (8.72) holds, let $\boldsymbol{\theta} : [0,\infty) \longrightarrow \mathbb{R}^d$ be a smooth map, and take $E_s^{\boldsymbol{\theta}}(t,\mathbf{W},\mathfrak{P})$ as in (10.40). Then, for each $\mathfrak{f} \in \mathcal{O}(M)$ and $s \in [0,1]$, $t \in [0,\infty) \longmapsto E_s^{\boldsymbol{\theta}}(t,\mathbf{W},\mathfrak{P}) \in (0,\infty)$ is a $\widehat{\mathbb{P}_{\mathfrak{f}}^{\boldsymbol{\theta}}}$-martingale. Thus, for each $s \in [0,1]$, there exists a unique probability measure $\mathbb{Q}_{s,\mathfrak{f}}^{\boldsymbol{\theta}}$ on $\mathcal{W}(\mathbb{R}^d) \times \mathcal{P}(\mathcal{O}(M))$ such that, for each $T \in (0,\infty)$,*

$$\mathbb{Q}_{s,\mathfrak{f}}^{\boldsymbol{\theta}}(C) = \mathbb{E}^{\widehat{\mathbb{P}_{\mathfrak{f}}^{\boldsymbol{\theta}}}}\left[E_s^{\boldsymbol{\theta}}(T),\, (\mathbf{W}_s,\mathfrak{P}_s) \in C\right] \quad \text{for } C \in \mathcal{F}_T.$$

Moreover, for each $\alpha \in (1,\infty)$,

$$\mathbb{E}^{\widehat{\mathbb{P}_{\mathfrak{f}}^{\boldsymbol{\theta}}}}\left[\sup_{t\in[0,T]} E_s^{\boldsymbol{\theta}}(t)^\alpha\right]^{\frac{1}{\alpha}} \le \frac{\alpha}{\alpha-1}\exp\left[(\alpha-1)s^2\int_0^T e^{-2\gamma t}|\dot{\boldsymbol{\theta}}(t)|^2\,dt\right]. \tag{10.47}$$

In particular,

$$s^2\int_0^\infty e^{-2\gamma t}|\dot{\boldsymbol{\theta}}(t)|^2\,dt < \infty \implies \mathbb{Q}_{s,\mathfrak{f}}^{\boldsymbol{\theta}} \ll (\mathbf{W}_s,\mathfrak{P}_s)_*\widehat{\mathbb{P}_{\mathfrak{f}}^{\boldsymbol{\theta}}}.$$

PROOF: Define $g_{s,\alpha} : \mathcal{W}^{(1)}(\mathbb{R}) \longrightarrow (0,\infty)$ so that

$$g_{s,\alpha}(\rho) = \exp\left(-\alpha\int_0^s \rho_\sigma\,d\sigma\right). \tag{10.48}$$

Then, for $1 \le k \le d$,

$$\widetilde{\mathfrak{X}_k(t,\boldsymbol{\theta})}_{(\mathfrak{f},a,\mathbf{v},j,\rho)} g_{s,\alpha} = -\alpha\left(\int_0^s \big(j_\sigma\dot{\boldsymbol{\theta}}(t), O_\sigma(a)\mathbf{e}_k\big)_{\mathbb{R}^d}\,d\sigma\right) g_{s,\alpha}(\rho), \tag{10.49}$$

and so (cf. (10.10), (10.19), and the proof of Corollary 10.28)

$$\begin{aligned}
\frac{\widetilde{\mathfrak{X}_k(t,\boldsymbol{\theta})}_{(\mathfrak{f},a,\mathbf{v},j,\rho)} \circ \widetilde{\mathfrak{X}_k(t,\boldsymbol{\theta})} g_{s,\alpha}}{g_{s,\alpha}(\rho)} &= \alpha^2\left(\int_0^s \big(j_\sigma\dot{\boldsymbol{\theta}}(t), O_\sigma(a)\mathbf{e}_k\big)_{\mathbb{R}^d}\,d\sigma\right)^2 \\
+\alpha\left(\iint\limits_{0\le\sigma_1\le\sigma_2\le s} \Big(j_{\sigma_2}\dot{\boldsymbol{\theta}}(t), O_{\sigma_1,\sigma_2}(a)\Omega_k(t,\boldsymbol{\theta})_{\sigma_1}(\mathfrak{f},a,\mathbf{v},j)O_{\sigma_1}(a)\mathbf{e}_k\Big)_{\mathbb{R}^d}\,d\sigma_1 d\sigma_2\right)&.
\end{aligned}$$

Next observe that

$$\begin{aligned}
\sum_{k=1}^d\left(\int_0^s \big(j_\sigma\dot{\boldsymbol{\theta}}(t), O_\sigma(a)\mathbf{e}_k\big)_{\mathbb{R}^d}\,d\sigma\right)^2 &= \left|\int_0^s O_\sigma(a)^\top j_\sigma\dot{\boldsymbol{\theta}}(t)\,d\sigma\right|^2 \\
&= 2\iint\limits_{0\le\sigma_1\le\sigma_2\le s}\Big(O_{\sigma_2}(a)^\top j_{\sigma_2}\dot{\boldsymbol{\theta}}(t), O_{\sigma_1}^\top(a)j_{\sigma_1}\dot{\boldsymbol{\theta}}(t)\Big)_{\mathbb{R}^d}\,d\sigma_1 d\sigma_2,
\end{aligned}$$

while

$$\sum_{k=1}^{d} \iint\limits_{0\le\sigma_1\le\sigma_2\le s} \Big(j_{\sigma_2}\dot{\boldsymbol{\theta}}(t), O_{\sigma_1,\sigma_2}(a)\Omega_k(t,\boldsymbol{\theta})_{\sigma_1}(\mathfrak{f},a,\mathrm{v},j)O_{\sigma_1}(a)\mathbf{e}_k\Big)_{\mathbb{R}^d}\,d\sigma_1 d\sigma_2$$
$$= -\iint\limits_{0\le\sigma_1\le\sigma_2\le s} \Big(O_{\sigma_2}(a)^\top j_{\sigma_2}\dot{\boldsymbol{\theta}}(t), O_{\sigma_1}(a)^\top \mathfrak{R}(t,\boldsymbol{\theta})_{\sigma_1}(\mathfrak{f},a,\mathrm{v},j)j_{\sigma_1}\boldsymbol{\theta}(t)\Big)_{\mathbb{R}^d}\,d\sigma_1 d\sigma_2.$$

Hence, after we put these together and do a minor re-arrangement, we arrive at

$$\sum_{k=1}^{d} \frac{\widetilde{\mathfrak{X}_k(t,\boldsymbol{\theta})}_{(\mathfrak{f},a,\mathbf{v},j,\rho)} \circ \widetilde{\mathfrak{X}_k(t,\boldsymbol{\theta})}g_{s,\alpha}}{g_{s,\alpha}(\rho)}$$
$$= \alpha(\alpha-1)\left|\int_0^s O_\sigma(a)^\top j_\sigma\dot{\boldsymbol{\theta}}(t)\,d\sigma\right|^2 + 2\alpha\int_0^s \Big(j_\sigma\dot{\boldsymbol{\theta}}(t), \mathbf{H}(t,\boldsymbol{\theta})(\mathfrak{f},a,\mathrm{v},j)\Big)_{\mathbb{R}^d}\,d\sigma.$$

But

$$\frac{\widetilde{\mathfrak{X}_0(t,\boldsymbol{\theta})}_{(\mathfrak{f},a,\mathbf{v},j,\rho)}g_{s,\alpha}}{g_{s,\alpha}(\rho)} = -\int_0^s \Big(j_\sigma\dot{\boldsymbol{\theta}}(t), \mathbf{H}(t,\boldsymbol{\theta})(\mathfrak{f},a,\mathrm{v},j)\Big)_{\mathbb{R}^d}\,d\sigma,$$

and therefore we have now shown that (cf. (10.44))

$$\widetilde{\mathcal{L}_t^{\boldsymbol{\theta}}}g_{s,\alpha} = \frac{\alpha(\alpha-1)}{2}\left|\int_0^s O_\sigma(a)^\top j_\sigma\dot{\boldsymbol{\theta}}(t)\,d\sigma\right|^2 g_{s,\alpha}. \tag{10.50}$$

Starting from the preceeding, we next apply Theorem 10.45 and the usual argument based on Doob's Stopping Time Theorem to see that if

$$\zeta_R = \inf\left\{t\ge 0 : \left|\int_0^s \rho_\sigma(t)\,d\sigma\right| \ge R\right\},$$

then

$$E_s^{\boldsymbol{\theta}}(t\wedge\zeta_R)^\alpha - \frac{\alpha(\alpha-1)}{2}\int_0^{t\wedge\zeta_R}\left|\int_0^s O_\sigma\big(A(\tau)\big)^\top J_\sigma(\tau)\dot{\boldsymbol{\theta}}(\tau)\,d\sigma\right|^2 E_s^{\boldsymbol{\theta}}(\tau)^\alpha\,d\tau$$

is a $\widehat{\mathbb{P}_{\mathfrak{f}}^{\boldsymbol{\theta}}}$-martingale. Thus, after using Gronwall's inequality, the estimate in (8.74), and Doob's inequality, we obtain the estimate in (10.47). In addition, when $\alpha = 1$, we have proved that $E_s^{\boldsymbol{\theta}}(t\wedge\zeta_R)$ is a $\widehat{\mathbb{P}_{\mathfrak{f}}^{\boldsymbol{\theta}}}$-martingale for each $R\in(0,\infty)$, and by (10.47), we know that $\{E_s^{\boldsymbol{\theta}}(t\wedge\zeta_R) : R>0\}$ is uniformly $\widehat{\mathbb{P}_{\mathfrak{f}}^{\boldsymbol{\theta}}}$-integrable. Therefore we can now say that $E_s^{\boldsymbol{\theta}}(t)$ is a $\widehat{\mathbb{P}_{\mathfrak{f}}^{\boldsymbol{\theta}}}$-martingale.

Finally, the $\widehat{\mathbb{P}_{\mathfrak{f}}^{\boldsymbol{\theta}}}$-martingale property for $E_s^{\boldsymbol{\theta}}(t)$ is equivalent to the assertion that $\mathbb{Q}_{s,\mathfrak{f}}^{\boldsymbol{\theta}}$ is consistently defined on $\{\mathcal{F}_t : t\ge 0\}$. Thus the existence of $\mathbb{Q}_{s,\mathfrak{f}}^{\boldsymbol{\theta}}$ on $\mathcal{B}_{\mathfrak{W}(\mathbb{R}^d)\times\mathcal{P}(\mathcal{O}(M))} = \bigvee_{t\ge 0}\mathcal{F}_t$ reduces to an elementary application of Kolmogorov's Consistency Theorem. $\square$

10.51 THEOREM. *Let everything be as in Lemma 10.46, and (cf. (10.32)) set* $\mathfrak{f}_s = \mathfrak{q}\big(\mathfrak{f}, \boldsymbol{\theta}(0)\big)_s$. *Then, for each* $s \in [0,1]$, $\mathbb{Q}^{\boldsymbol{\theta}}_{s,\mathfrak{f}}$ *is the distribution of (cf. (10.23))* $\mathbf{w} \rightsquigarrow \big(\mathbf{w}, \mathfrak{p}(\,\cdot\,, \mathfrak{f}_s, \mathbf{w})\big)$ *under* $\mu_{\mathbb{R}^d}$. *Equivalently, for each* $T \in (0,\infty)$ *and* $\Phi : \mathfrak{W}(\mathbb{R}^d) \times \mathcal{P}\big(\mathcal{O}(M)\big) \longrightarrow \mathbb{R}$ *which is bounded and* $\mathcal{F}_T$*-measurable,*

$$\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\Phi \circ \big(\mathbf{w}, \mathfrak{p}(\,\cdot\,, \mathfrak{f}_s, \mathbf{w})\big)\Big] = \mathbb{E}^{\widehat{\mathbb{P}^{\boldsymbol{\theta}}_{\mathfrak{f}}}}\Big[E^{\boldsymbol{\theta}}_s(T)\Phi \circ \big(\mathbf{W}_s, \mathfrak{P}_s\big)\Big] \tag{10.52}$$

for each $s \in [0,1]$, *and so (cf. (8.77) and (10.40))*

$$\begin{aligned} &\frac{d}{ds}\left(\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[\Phi \circ \big(\mathbf{w}, \mathfrak{p}(\,\cdot\,, \mathfrak{f}_s, \mathbf{w})\big)\Big] - \mathbb{E}^{\widehat{\mathbb{P}^{\boldsymbol{\theta}}_{\mathfrak{f}}}}\Big[\Phi \circ \big(\mathbf{W}_s, \mathfrak{P}_s\big)\Big]\right)\Bigg|_{s=0} \\ &\qquad = -\mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[B_{\boldsymbol{\theta}}(T,\mathfrak{f},\mathbf{w})\Phi \circ \big(\mathbf{w}, \mathfrak{p}(\,\cdot\,, \mathfrak{f}, \mathbf{w})\big)\Big] \\ &\text{where } B_{\boldsymbol{\theta}}(T,\mathfrak{f},\mathbf{w}) \equiv \big[B(\boldsymbol{\theta})\big]\big(T, \mathbf{w}, \mathfrak{p}(\,\cdot\,, \mathfrak{f}, \mathbf{w})\big). \end{aligned} \tag{10.53}$$

PROOF: In order to prove the first part, we have to check that

$$\varphi\big(\mathbf{W}_s(t), \mathfrak{P}_s(t)\big) - \int_0^t \frac{1}{2}\sum_{k=1}^d \big(\partial_{\mathbf{e}_k} + \mathfrak{E}(\mathbf{e}_k)\big)^2 \varphi\big(\mathbf{W}_s(\tau), \mathfrak{P}_s(\tau)\big)\,d\tau$$

is a $\mathbb{Q}^{\boldsymbol{\theta}}_{s,\mathfrak{f}}$-martingale for each $\varphi \in C^\infty_c\big(\mathbb{R}^d \times \mathcal{O}(M);\mathbb{R}\big)$. But, because $t \rightsquigarrow E^{\boldsymbol{\theta}}_s(t)$ is a $\widehat{\mathbb{P}^{\boldsymbol{\theta}}_{\mathfrak{f}}}$-martingale, this is equivalent to checking that

$$E^{\boldsymbol{\theta}}_s(t)\varphi\big(\mathbf{W}_s(t), \mathfrak{P}_s(t)\big) - \int_0^t E^{\boldsymbol{\theta}}_s(\tau)\frac{1}{2}\sum_{k=1}^d \big(\partial_{\mathbf{e}_k} + \mathfrak{E}(\mathbf{e}_k)\big)^2 \varphi\big(\mathbf{W}_s(\tau), \mathfrak{P}_s(\tau)\big)\,d\tau$$

is a $\widehat{\mathbb{P}^{\boldsymbol{\theta}}_{\mathfrak{f}}}$-martingale. To this end, take $g_{s,1}$ as in the proof of Lemma 10.46, and note that, by (10.50), $\widetilde{\mathcal{L}^{\boldsymbol{\theta}}_t} g_{s,1} \equiv 0$. Hence, after combining this with (10.31) and (10.49), we find that, at the point

$$\Big(\mathbf{W}_s(\tau), \mathfrak{F}(\tau,\boldsymbol{\theta})_s\big(\mathfrak{p}(\tau), A(\tau), \mathbf{W}(\tau), J(\tau)\big), \rho(\tau)\Big),$$

one has

$$\widetilde{\mathcal{L}^{\boldsymbol{\theta}}_t}(g_{s,1}\varphi) = g_{s,1}\big(\rho(\tau)\big)\frac{1}{2}\sum_{k=1}^d\Big(\partial_{\mathbf{e}_k} + \mathfrak{E}(\mathbf{e}_k)\Big)^2\varphi\big(\mathbf{W}_s(\tau), \mathfrak{P}_s(\tau)\big),$$

which proves that

$$\begin{aligned} &g_{s,1}\big(\rho(t)\big)\varphi\big(\mathbf{W}_s(t), \mathfrak{P}_s(t)\big) \\ &\qquad - \int_0^t g_{s,1}\big(\rho(\tau)\big)\frac{1}{2}\sum_{k=1}^d \big(\partial_{\mathbf{e}_k} + \mathfrak{E}(\mathbf{e}_k)\big)^2 \varphi\big(\mathbf{W}_s(\tau), \mathfrak{P}_s(\tau)\big)\,d\tau \end{aligned}$$

is a $\widehat{\mathbb{P}_{\mathfrak{f}}^{\boldsymbol{\theta}}}$-martingale. Hence, because $E_s^{\boldsymbol{\theta}}(t) = g_{s,1}(\rho(t))$ $\widetilde{\mathbb{P}_{\mathfrak{f}}^{\boldsymbol{\theta}}}$-almost surely, we have now checked the first assertion.

To complete the proof from here, observe that (10.52) is nothing but a restatement of the first assertion and that (10.53) follows by differentiating (10.52) at $s = 0$ and applying the estimates in (8.79) and (10.47) to justify bringing the derivative under the integral. □

As a consequence of (10.53), applied to $\Phi(\mathbf{w}, \mathfrak{p}) = f \circ \pi(\mathfrak{p}(T))$, and the estimate in (8.65), we have the following statement of Bismut's formula (cf. (8.80)).

10.54 Corollary. *If $f \in C(M;\mathbb{R})$ satisfies*

$$\lim_{r\to\infty} \sup\left\{\frac{\log(1+|f(x)|)}{\operatorname{dist}^M(x,x_0)^2} : \operatorname{dist}^M(x,x_0) \ge r\right\} = 0$$

for some reference point $x_0 \in M$, then, for all $(T, \mathfrak{f}) \in [0,\infty) \times \mathcal{O}(M)$, (cf. (8.77))

$$\sup_{t\in[0,T]} \left|B_{\boldsymbol{\theta}}(t,\mathfrak{f},\mathbf{w}) f\circ\pi(\mathfrak{p}(t,\mathfrak{f},\mathbf{w}))\right| \in L^1(\mu_{\mathbb{R}^d})$$

and

$$\mathfrak{E}(\boldsymbol{\theta}(T) - \boldsymbol{\theta}(0))_{\mathfrak{f}}\left[\mathbf{P}_T^M f\right]\circ\pi = \mathbb{E}^{\mu_{\mathbb{R}^d}}\Big[B_{\boldsymbol{\theta}}(T,\mathfrak{f},\mathbf{w}) f\circ\pi(\mathfrak{p}(T,\mathfrak{f},\mathbf{w}))\Big], \tag{10.55}$$

which is the same statement as (8.80).

10.5 Second Derivatives

In this penultimate section, we will examine what happens when, by taking two derivatives of (10.52), one uses the considerations in § 10.4 to get information about the Hessian of $\mathbf{P}_T^M f$.

10.5.1. Derivative of the Bismut Factor. As soon as one differentiates (10.52) twice, it becomes necessary to deal with the quantity

$$[B(\boldsymbol{\theta})]_s'(T, \mathbf{W}, \mathfrak{P}) \equiv \frac{d}{ds}[B(\boldsymbol{\theta})](T, \mathbf{W}_s, \mathfrak{P}_s). \tag{10.56}$$

First note that, after integration by parts, $[B(\boldsymbol{\theta})](T, \mathbf{W}_s, \mathfrak{P}_s)$ can be written as

$$\left(J_s(T)\dot{\boldsymbol{\theta}}(T), \mathbf{W}_s(T)\right)_{\mathbb{R}^d} + \int_0^T \left(\tfrac{1}{2}\mathfrak{R}_{\mathfrak{P}_s(t)} J_s(t)\dot{\boldsymbol{\theta}}(t) - J_s(t)\ddot{\boldsymbol{\theta}}(t), \mathbf{W}_s(t)\right)_{\mathbb{R}^d} dt,$$

and so there is no question about the existence of $[B(\boldsymbol{\theta})]_s'(T, \mathbf{W}, \mathfrak{P})$. On the other hand, it should be apparent that the size of $[B(\boldsymbol{\theta})]_s'(T, \mathbf{W}, \mathfrak{P})$ cannot be controlled in terms of lower bounds on the Ricci curvature alone. Indeed, learning how to estimate the size of $[B(\boldsymbol{\theta})]_s'(T, \mathbf{W}, \mathfrak{P})$ is the goal of this subsection, and for this purpose we will need to use the following lemma.

10.57 LEMMA. *Suppose that the time-dependent operator $t \rightsquigarrow \mathcal{L}_t$ is given by (3.20) for some time-dependent vector fields $t \rightsquigarrow X_k(t)$, $0 \le k \le d$, on some space $\mathcal{M}$ which has a differentiable structure. Given a smooth, real-valued function ψ on $\mathcal{M}$, define $Y_\psi : [0,\infty) \times \mathcal{P}(\mathcal{M}) \longrightarrow \mathbb{R}$ by*

$$Y_\psi(t,p) = \psi\big(p(t)\big) - \int_0^t \mathcal{L}_\tau \psi\big(p(\tau)\big)\, d\tau.$$

Finally, suppose that $\mathfrak{D}$ is an algebra of smooth functions on $\mathcal{M}$ with the property that $\psi \in \mathfrak{D} \implies e^\psi \in \mathfrak{D}$ and that $\mathbb{P}$ is a probability measure on $\mathcal{P}(\mathcal{M})$ such that $Y_\psi(t \wedge \Sigma, p)$ is a $\mathbb{P}$-martingale for some stopping time Σ and every $\psi \in \mathfrak{D}$. Then, for each $\psi \in \mathfrak{D}$,

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}&\left[\exp\left(\sup_{t\in[0,T\wedge\Sigma]}\left|Y_\psi(t) - \psi\big(p(0)\big) - \int_0^t \mathcal{L}_\tau\psi\big(p(\tau)\big)\,d\tau\right|\right)\right] \\ &\le 8\mathbb{E}^{\mathbb{P}}\left[\exp\left(2\sum_{k=1}^d \int_0^{T\wedge\Sigma} \big(X_k(t)_{p(t)}\psi\big)^2\,dt\right)\right]^{\frac{1}{2}}.\end{aligned}$$

Moreover, if $\psi \in \mathfrak{D}$ satisfies

$$\sup_{p\in\mathcal{P}(\mathcal{M})} \sum_{k=1}^d \int_0^{T\wedge\Sigma} \big(X_k(t)_{p(t)}\psi\big)^2\,dt \le A < \infty,$$

then

$$\mathbb{E}^{\mathbb{P}}\left[\exp\left(\sup_{t\in[0,T\wedge\Sigma]} \frac{\big(Y_\psi(t,p) - \psi(p(0))\big)^2}{A}\right)\right] \le 12.$$

PROOF: First note that

$$\mathcal{L}_t e^\psi = \left(\mathcal{L}_t\psi + \frac{1}{2}\sum_{k=1}^d \big(X_k(t)\psi\big)^2\right) e^\psi,$$

and set

$$\Xi(t,p) = \int_0^t \left(\mathcal{L}_\tau\psi\big(p(\tau)\big) + \frac{1}{2}\sum_{k=1}^d \big(X_k(\tau)_{p(\tau)}\psi\big)^2\right)\,d\tau.$$

Without loss in generality, we may and will assume that Σ is taken so that $|\dot\Xi(t\wedge\Sigma,p)|$ is uniformly bounded as (t,p) ranges over $t \in [0,\infty) \times \mathcal{P}(\mathcal{M})$. Thus, since

$$e^{\psi(t\wedge\Sigma)} - \int_0^{t\wedge\Sigma} \left[\mathcal{L}_\tau e^\psi\right]\big(p(\tau)\big)\,d\tau$$

is a $\mathbb{P}$-martingale, we can apply Lemma 2.41 to see that

$$\begin{aligned}\exp\Big(\psi\big(p(t\wedge\Sigma)\big)-\Xi\big(t\wedge\Sigma,p\big)\Big)&-e^{-\Xi(t\wedge\Sigma,p)}\int_0^{t\wedge\Sigma}e^{\psi(p(\tau))}\,d\Xi(\tau,p)\\&+\int_0^{t\wedge\Sigma}e^{\psi(p(\tau))-\Xi(\tau,p)}\,d\Xi(\tau,p)\\&-\int_0^{t\wedge\Sigma}e^{-\Xi(\tau,p)}\left(\int_0^\tau e^{\psi(p(\sigma))}\,d\Xi(\sigma,p)\right)d\Xi(\tau,p)\end{aligned}$$

is also a $\mathbb{P}$-martingale. But, after integrating the last term by parts, we see that all but the first term cancel. Hence, we have now proved that, for each $\psi\in\mathfrak{D}$,

$$\begin{gathered}E_\psi(t,p)\equiv\exp\Big(Y_\psi\big(p(t\wedge\Sigma)\big)-\psi\big(p(0)\big)-\tfrac12 S_\psi(t\wedge\Sigma,p)\Big)\\ \text{with } S_\psi(t,p)\equiv\sum_{k=1}^d\int_0^t\big(X_k(\tau)_{p(\tau)}\psi\big)^2\,d\tau\end{gathered}$$

is a $\mathbb{P}$-martingale.

To complete the proof from here, we begin by observing that, for each $\lambda\in\mathbb{R}$ and $\psi\in\mathfrak{D}$, $\exp\big(\lambda\bar{Y}_\psi(t,p)\big)$ is a non-negative $\mathbb{P}$-submartingale when $\bar{Y}_\psi(t,p)\equiv Y_\psi(t\wedge\Sigma,p)-\psi\big(p(0)\big)$. Hence, by Doob's inequality,

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\left[\sup_{t\in[0,T]}\exp\big(\lambda\bar{Y}_\psi(t,p)\big)\right]&=\mathbb{E}^{\mathbb{P}}\left[\sup_{t\in[0,T]}\exp\left(\frac{\lambda}{2}\bar{Y}_\psi(t,p)\right)^2\right]\\&\le 4\mathbb{E}^{\mathbb{P}}\left[\exp\left(\frac{\lambda}{2}\bar{Y}_\psi(T,p)\right)^2\right]=4\mathbb{E}^{\mathbb{P}}\Big[\exp\big(\lambda\bar{Y}_\psi(T,p)\big)\Big],\end{aligned}$$

from which it is clear that

$$(*)\qquad \mathbb{E}^{P}\left[e^{\lambda\bar{Y}^*_\psi(T,p)}\right]\le 4\mathbb{E}^{\mathbb{P}}\left[e^{\lambda\bar{Y}_\psi(T,p)}+e^{-\lambda\bar{Y}_\psi(T,p)}\right],$$

where $\bar{Y}^*_\psi(T,p)\equiv\sup_{t\in[0,T]}|\bar{Y}_\psi(t,p)|$. Next, observe that

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\left[e^{\pm\bar{Y}_\psi(T,p)}\right]&=\mathbb{E}^{\mathbb{P}}\left[e^{\pm\bar{Y}_\psi(T,p)-S_\psi(T,p)}e^{S_\psi(T,p)}\right]\\&\le\mathbb{E}^{\mathbb{P}}\big[E_{\pm2\psi}(T,p)\big]^{\frac12}\mathbb{E}^{\mathbb{P}}\left[e^{2S_\psi(T,p)}\right]^{\frac12}=\mathbb{E}^{\mathbb{P}}\left[e^{2S_\psi(T,p)}\right]^{\frac12},\end{aligned}$$

which, in conjunction with (*), completes the proof of the first estimate. As for the second estimate, when $S_\psi(T,p)\le A$, we again apply (*) to see that

$$\mathbb{E}^{\mathbb{P}}\left[e^{\lambda\bar{Y}^*_\psi(T,p)}\right]\le 4\mathbb{E}^{\mathbb{P}}\big[E_{\lambda\psi}(T,p)+E_{-\lambda\psi}(T,p)\big]\exp\left(\frac{\lambda^2A}{2}\right)=8\exp\left(\frac{\lambda^2A}{2}\right)$$

for all $\lambda \in \mathbb{R}$. Hence, after multiplying through by $e^{-\lambda^2 A}$ and integrating with respect to λ, we arrive at the second estimate. $\square$

We now want to apply Lemma 10.57 to the problem of estimating $[B(\boldsymbol{\theta})]'_s$. To this end, recall (cf. § 10.4.1) that $[B(\boldsymbol{\theta})](\,\cdot\,, \mathfrak{P}, \mathbf{W})$ has the same distribution under $\widehat{\mathbb{P}_{\mathfrak{f}}^{\boldsymbol{\theta}}}$ as ρ has under $\widetilde{\mathbb{P}_{\mathfrak{f}}^{\boldsymbol{\theta}}}$. Hence, we begin by examining $\widetilde{\mathfrak{X}_k(t,\boldsymbol{\theta})}\rho_s$ and $\widetilde{\mathcal{L}_t^{\boldsymbol{\theta}}}\rho_s$. But, (cf. (10.48))

$$\rho_s = -\frac{\partial^2 g_{\alpha,s}}{\partial s \partial \alpha}(\rho)\Big|_{\alpha=0},$$

and so, by (10.49) and (10.50), we have that

$$\begin{aligned}
&\widetilde{\mathfrak{X}_k(t,\boldsymbol{\theta})}_{(\mathfrak{f},a,\mathbf{v},\jmath,\rho)}\rho_s = \Big(\jmath_s\dot{\boldsymbol{\theta}}(t), O_s(a)\mathbf{e}_k\Big)_{\mathbb{R}^d}, \quad 1 \le k \le d \\
&\big[\widetilde{\mathcal{L}_t^{\boldsymbol{\theta}}}\rho_s\big](\mathfrak{f},a,\mathbf{v},\jmath,\rho) = \int_0^s \Big(O_s(a)^\top \jmath_s\dot{\boldsymbol{\theta}}(t), O_\sigma(a)^\top \jmath_\sigma\dot{\boldsymbol{\theta}}(t)\Big)_{\mathbb{R}^d}\,d\sigma.
\end{aligned} \tag{10.58}$$

Next, by taking one more derivative with respect of s, we obtain

$$\begin{aligned}
&\sum_{k=1}^d \Big(\widetilde{\mathfrak{X}_k(t,\boldsymbol{\theta})}_{(\mathfrak{f},a,\mathbf{v},\jmath,\rho)}\rho'_s\Big)^2 = \big|(\jmath'_s + a_s\jmath_s)\dot{\boldsymbol{\theta}}(t)\big|^2_{\mathbb{R}^d} \\
&\big[\widetilde{\mathcal{L}_t^{\boldsymbol{\theta}}}\rho'_s\big](\mathfrak{f},a,\mathbf{v},\jmath,\rho) \\
&\qquad = \big|\jmath_s\dot{\boldsymbol{\theta}}(t)\big|^2_{\mathbb{R}^d} + \int_0^s \Big(O_s(a)^\top\big(\jmath'_s + a_s\jmath_s\big)\dot{\boldsymbol{\theta}}(t), O_\sigma(a)^\top \jmath_\sigma\dot{\boldsymbol{\theta}}(t)\Big)_{\mathbb{R}^d}\,d\sigma.
\end{aligned} \tag{10.59}$$

As the preceding makes clear, before we can use (10.59) in Lemma 10.57, we must still come to terms with the size of J'_s and A_s under $\widehat{\mathbb{P}_{\mathfrak{f}}^{\boldsymbol{\theta}}}$. To understand J'_s, remember that, $\widehat{\mathbb{P}_{\mathfrak{f}}^{\boldsymbol{\theta}}}$-almost surely, $J_s(0) = I$ and (cf. (10.9) and (10.23)) $\dot{J}_s(t) + \frac{1}{2}\mathfrak{R}_{\mathfrak{P}_s(t)}J_s(t) = 0$. Thus, $\widehat{\mathbb{P}_{\mathfrak{f}}^{\boldsymbol{\theta}}}$-almost surely,

$$\frac{d}{dt}J'_s(t) + \tfrac{1}{2}\mathfrak{R}_{\mathfrak{P}_s(t)}J'_s(t) = -\tfrac{1}{2}\mathfrak{R}'_{\mathfrak{P}_s(t)}J_s(t),$$

where

$$\mathfrak{R}'_{\mathfrak{P}_s(t)} \equiv \frac{d}{ds}\mathfrak{R}_{\mathfrak{P}_s(t)} = \mathfrak{C}\big(J_s(t)\boldsymbol{\theta}(t)\big)_{\mathfrak{P}_s(t)}\mathfrak{R} + \lambda\big(A_s(t)\big)_{\mathfrak{P}_s(t)}\mathfrak{R}.$$

Hence, by (8.59) and (8.73),

$$\begin{aligned}
J'_s(t) = -\frac{1}{2}\mathfrak{J}_{\mathfrak{P}_s}(t)\int_0^t \mathfrak{J}_{\mathfrak{P}_s}(\tau)^{-1}\Big(&\mathfrak{C}\big(\mathfrak{J}_{\mathfrak{P}_s}(\tau)\boldsymbol{\theta}(\tau)\big)_{\mathfrak{P}_s(\tau)}\mathfrak{R} \\
&+ \big[A_s(\tau), \mathfrak{R}_{\mathfrak{P}_s(\tau)}\big]\Big)\mathfrak{J}_{\mathfrak{P}_s}(\tau)\,d\tau
\end{aligned} \tag{10.60}$$

$\widehat{\mathbb{P}^{\theta}_{\mathfrak{f}}}$-almost surely.

Turning to A_s, first observe that

$$\widehat{\mathfrak{X}_k(t,\theta)}_{(\mathfrak{f},a,\mathrm{v},j)} a_s = \Omega_{\mathfrak{F}_s(\mathfrak{f},a,j\theta(t))}\big(O_s(a)\mathbf{e}_k, j_s\theta(t)\big) \quad \text{for } 1 \le k \le d.$$

Thus, by very much the same reasoning as we used in the derivation of (10.50), we find that

$$\begin{aligned}\big[\widehat{\mathcal{L}^{\theta}_t} a_s\big](\mathfrak{f},a,\mathrm{v},j) = \int_0^s &\Omega_{\mathfrak{F}_s(\mathfrak{f},a,j\theta(t))}\big(O_{\sigma,s}(a)j_\sigma\dot{\theta}(t), j_s\theta(t)\big)\,d\sigma \\ &+ \frac{1}{2}\sum_{k=1}^d \mathfrak{C}\big(O_s(a)\mathbf{e}_k\big)_{\mathfrak{F}_s(\mathfrak{f},a,j\theta(t))}\Omega\big(O_s(a)\mathbf{e}_k, j_s\theta(t)\big).\end{aligned}$$

Next, observe that, by (R4) and the second Bianci identity (cf. (8.52)), for any $\boldsymbol{\xi}$, $\boldsymbol{\eta}$, and $\boldsymbol{\zeta}$ from $\mathbb{R}^d$:

$$\begin{aligned}\left(\sum_{k=1}^d \mathfrak{C}(O_s(a)\mathbf{e}_k)\Omega(O_s(a)\mathbf{e}_k,\boldsymbol{\xi})\boldsymbol{\eta},\boldsymbol{\zeta}\right)_{\mathbb{R}^d} &= \left(\sum_{k=1}^d \mathfrak{C}(\mathbf{e}_k)\Omega(\boldsymbol{\eta},\boldsymbol{\zeta})\mathbf{e}_k,\boldsymbol{\xi}\right)_{\mathbb{R}^d} \\ = -\left(\sum_{k=1}^d\Big(\mathfrak{C}(\boldsymbol{\eta})\Omega(\boldsymbol{\zeta},\mathbf{e}_k) + \mathfrak{C}(\boldsymbol{\zeta})\Omega(\mathbf{e}_k,\boldsymbol{\eta})\Big)\mathbf{e}_k,\boldsymbol{\xi}\right)_{\mathbb{R}^d} &= \big(\mathcal{A}(\boldsymbol{\xi})\boldsymbol{\eta},\boldsymbol{\zeta}\big)_{\mathbb{R}^d},\end{aligned}$$

where $\mathcal{A}_{\mathfrak{f}} \in \mathrm{Hom}(\mathbb{R}^d; o(d))$ is defined for $\mathfrak{f} \in \mathcal{O}(M)$ so that

$$\big(\mathcal{A}(\boldsymbol{\xi})\boldsymbol{\eta},\boldsymbol{\zeta}\big)_{\mathbb{R}^d} = \big(\mathfrak{C}(\boldsymbol{\zeta})\mathfrak{R}\boldsymbol{\eta} - \mathfrak{C}(\boldsymbol{\eta})\mathfrak{R}\boldsymbol{\zeta},\boldsymbol{\xi}\big)_{\mathbb{R}^d}. \tag{10.61}$$

Hence, we now know that, for any $\boldsymbol{\xi}$, $\boldsymbol{\eta} \in \mathbb{R}^d$,

$$\begin{aligned}&\sum_{k=1}^d\Big(\widehat{\mathfrak{X}_k(t,\theta)}_{(\mathfrak{f},a,\mathrm{v},j,\rho)}(a_s\boldsymbol{\xi},\boldsymbol{\eta})_{\mathbb{R}^d}\Big)^2 = \big|\Omega_{\mathfrak{f}}(\boldsymbol{\xi},\boldsymbol{\eta})j_s\theta(t)\big|^2_{\mathbb{R}^d} \\ &\big[\widehat{\mathcal{L}^{\theta}_t}(a_s\boldsymbol{\xi},\boldsymbol{\eta})_{\mathbb{R}^d}\big](\mathfrak{f},a,\mathrm{v},j,\rho) \\ &\qquad = \frac{1}{2}\Big(\mathcal{A}_{\mathfrak{F}_s(\mathfrak{f},a,j\theta(t))}\big(j_s\theta(t)\big)\boldsymbol{\xi},\boldsymbol{\eta}\Big)_{\mathbb{R}^d} \\ &\qquad\quad + \int_0^s\Big(\Omega_{\mathfrak{F}_s(\mathfrak{f},a,j\theta(t))}\big(O_{\sigma,s}(a)j_\sigma\dot{\theta}(t), j_s\theta(t)\big)\boldsymbol{\xi},\boldsymbol{\eta}\Big)_{\mathbb{R}^d}\,d\sigma,\end{aligned} \tag{10.62}$$

where, in order to get the first line, we have used (8.54) together with (R4) to see that

$$\sum_{k=1}^d\Big(\Omega(O_s(a)\mathbf{e}_k, j_s\theta(t))\boldsymbol{\xi},\boldsymbol{\eta}\Big)^2_{\mathbb{R}^d} = \sum_{k=1}^d\Big(\Omega(\boldsymbol{\xi},\boldsymbol{\eta})\mathbf{e}_k, j_s\theta(t)\Big)^2_{\mathbb{R}^d} = \big|\Omega(\boldsymbol{\xi},\boldsymbol{\eta})j_s\theta(t)\big|^2_{\mathbb{R}^d}.$$

10.63 LEMMA. *Assume that (8.72) holds and, for $\boldsymbol{\eta} \in C([0,\infty);\mathbb{R}^d)$, define*

$$T \in [0,\infty) \longmapsto \Gamma_\gamma(T,\boldsymbol{\eta}) \equiv \int_0^T \left(e^{-\gamma t}|\boldsymbol{\eta}(t)|_{\mathbb{R}^d}\right)^2 dt \in \mathbb{R}. \tag{10.64}$$

Then, for each $(T,\mathfrak{f}) \in (0,\infty) \times \mathcal{O}(M)$ and $s \in [0,1]$,

$$\mathbb{E}^{\widehat{\mathbb{P}^{\boldsymbol{\theta}}_{\mathfrak{f}}}}\left[\exp\left(\frac{[B(\boldsymbol{\theta})](T,\mathbf{W}_s,\mathfrak{P}_s)^2}{(1+s)\Gamma_\gamma(T,\dot{\boldsymbol{\theta}})}\right)\right] \le 12\exp\big(s\Gamma_\gamma(T,\dot{\boldsymbol{\theta}})\big). \tag{10.65}$$

Next, for $x \in M$ and $R > 0$, set

$$K_\Omega(x,R) = \sup\Big\{\|\Omega_{\mathfrak{f}}(\boldsymbol{\xi},\boldsymbol{\eta})\|_{\text{op}} : \operatorname{dist}^M(\pi\mathfrak{f},x) \le R \text{ and } \boldsymbol{\xi},\,\boldsymbol{\eta} \in \mathbf{S}^{d-1}\Big\}.$$

Then, for all $(T,x) \in (0,\infty) \times M$, $\mathfrak{f} \in \pi^{-1}(x)$, and $R \in (0,\infty)$,

$$\begin{aligned}\mathbb{E}^{\widehat{\mathbb{P}^{\boldsymbol{\theta}}_{\mathfrak{f}}}}&\left[\exp\left(\sup_{t\in[0,T\wedge\zeta_R]}\frac{\|\overline{A_s}(t)\|^2_{\text{op}}}{(1+s)d^2K_\Omega\big(x,R+r_\gamma(T,\boldsymbol{\theta})\big)\Gamma_\gamma(T,\boldsymbol{\theta})}\right)\right]\\ &\le 12\exp\Big(sK_\Omega\big(x,R+r_\gamma(T,\boldsymbol{\theta})\big)\sqrt{\Gamma_\gamma(T,\boldsymbol{\theta})\Gamma_\gamma(T,\dot{\boldsymbol{\theta}})}\Big),\end{aligned} \tag{10.66}$$

where (cf. (10.61))

$$\begin{aligned}\zeta_R &= \inf\big\{t \ge 0 : \operatorname{dist}^M\big(\pi\mathrm{p}(t),\pi\mathrm{p}(0)\big) \ge R\big\},\\ \overline{A_s}(t) &= A_s(t) - \int_0^t \tfrac{1}{2}A_s\big(J_s(\tau)\boldsymbol{\theta}(\tau)\big)\,d\tau,\\ r_\gamma(T,\boldsymbol{\theta}) &= e^{-\gamma^- T}\sup_{t\in[0,T]}|\boldsymbol{\theta}(t)|_{\mathbb{R}^d}.\end{aligned} \tag{10.67}$$

Finally, for each $(T,\mathfrak{f}) \in (0,\infty) \times \mathcal{O}(M)$, $\lambda \in (0,\infty)$, and $R \in (0,\infty)$,

$$\begin{aligned}\mathbb{E}^{\widehat{\mathbb{P}^{\boldsymbol{\theta}}_{\mathfrak{f}}}}&\left[\exp\left(\lambda\sup_{t\in[0,T\wedge\zeta_R]}\big|[B(\boldsymbol{\theta})]'_s(t,\mathbf{W},\mathfrak{P})\big|_{\mathbb{R}^d}\right)\right]\\ &\le 8e^{(\lambda+2^{-1}s)\Gamma_\gamma(T,\dot{\boldsymbol{\theta}})}\mathbb{E}^{\widehat{\mathbb{P}^{\boldsymbol{\theta}}_{\mathfrak{f}}}}\left[\exp\left(2^{-1}\big((1+s)\lambda\big)^2\int_0^{T\wedge\zeta_R}\big|\hat{J}_s(t)\dot{\boldsymbol{\theta}}(t)\big|^2_{\mathbb{R}^d}\,dt\right)\right],\end{aligned} \tag{10.68}$$

where

$$\hat{J}_s(t) \equiv J'_s(t) + A_s(t)J_s(t).$$

PROOF: In proving (10.65) and (10.68), remember that the distributions of

$$[B(\boldsymbol{\theta})](\,\cdot\,,\mathbf{W}_s,\mathfrak{P}_s) \quad\text{and}\quad [B(\boldsymbol{\theta})]'_s(\,\cdot\,,\mathbf{W},\mathfrak{P}) \text{ under } \widehat{\mathbb{P}^{\boldsymbol{\theta}}_{\mathfrak{f}}}$$

are the same as those of $\rho_s(\,\cdot\,)$ and $\rho'_s(\,\cdot\,)$, respectively, under $\widetilde{\mathbb{P}^{\theta}_{\mathfrak{f}}}$. In particular, (10.65) follows easily when we feed the calculations in (10.58) into the second part of Lemma 10.57. Similarly, (10.66) follows from the calculations in (10.62) once one recalls (cf. § 10.3.2) that

$$\operatorname{dist}^M\bigl(\pi\circ\mathfrak{P}_s(t),\pi\circ\mathfrak{p}(t)\bigr)\le r_\gamma(t,\boldsymbol{\theta})\quad \widehat{\mathbb{P}^{\theta}_{\mathfrak{f}}}\text{-almost surely.}$$

Thus, all that remains is to prove (10.68). But we know that (cf. the notation used in Lemma 10.57)

$$\begin{aligned}
&\mathbb{E}^{\widetilde{\mathbb{P}^{\theta}_{\mathfrak{f}}}}\left[\exp\left(\lambda\sup_{t\in[0,T\wedge\zeta_R]}\bigl|\rho'_s(t)\bigr|_{\mathbb{R}^d}\right)\right]\\
&\quad\le \mathbb{E}^{\widetilde{\mathbb{P}^{\theta}_{\mathfrak{f}}}}\left[\exp\left(\lambda\sup_{t\in[0,T\wedge\zeta_R]}\bigl|Y_{\rho'_s}(t)\bigr|_{\mathbb{R}^d}\right)\exp\left(\lambda\int_0^{T\wedge\zeta_R}\bigl|\widetilde{\mathcal{L}^{\theta}_t}\rho'_s(t)\bigr|\,dt\right)\right]\\
&\quad\le e^{(\lambda+2^{-1}s)\Gamma_\gamma(T,\dot{\boldsymbol{\theta}})}\,\mathbb{E}^{\widetilde{\mathbb{P}^{\theta}_{\mathfrak{f}}}}\left[\exp\left(\lambda\sup_{t\in[0,T\wedge\zeta_R]}\bigl|Y_{\rho'_s}(t)\bigr|_{\mathbb{R}^d}\right)\right.\\
&\qquad\qquad\left.\times\exp\left(2^{-1}s\lambda^2\int_0^{T\wedge\zeta_R}\bigl|\hat{J}_s(t)\dot{\boldsymbol{\theta}}(t)\bigr|^2_{\mathbb{R}^d}\,dt\right)\right],
\end{aligned}$$

where, in the passage to the last inequality, we have used (cf. (10.59))

$$\lambda\int_0^{T\wedge\zeta_R}\bigl|\widetilde{\mathcal{L}^{\theta}_t}\rho'_s(t)\bigr|\,dt\le(\lambda+2^{-1}s)\Gamma_\gamma(T,\dot{\boldsymbol{\theta}})+2^{-1}s\lambda^2\int_0^{T\wedge\zeta_R}\bigl|\hat{J}_s(t)\dot{\boldsymbol{\theta}}(t)\bigr|^2_{\mathbb{R}^d}\,dt.$$

We next use the first line of (10.59) and apply Hölder's inequality and the first part of Lemma 10.57 to see that the left hand side of (10.68) is dominated by

$$\begin{aligned}
&8e^{(\lambda+2^{-1}s)\Gamma_\gamma(T,\dot{\boldsymbol{\theta}})}\,\mathbb{E}^{\widehat{\mathbb{P}^{\theta}_{\mathfrak{f}}}}\left[\exp\left(2\bigl((1+s)\lambda\bigr)^2\int_0^{T\wedge\zeta_R}\bigl|\hat{J}_s(t)\dot{\boldsymbol{\theta}}(t)\bigr|^2_{\mathbb{R}^d}\,dt\right)\right]^{\frac{1}{2(1+s)}}\\
&\qquad\times\mathbb{E}^{\widehat{\mathbb{P}^{\theta}_{\mathfrak{f}}}}\left[\exp\left(2^{-1}(1+s)\lambda^2\int_0^{T\wedge\zeta_R}\bigl|\hat{J}_s(t)\dot{\boldsymbol{\theta}}(t)\bigr|^2_{\mathbb{R}^d}\,dt\right)\right]^{\frac{s}{1+s}}\\
&\quad\le 8e^{(\lambda+2^{-1}s)\Gamma_\gamma(T,\dot{\boldsymbol{\theta}})}\,\mathbb{E}^{\widehat{\mathbb{P}^{\theta}_{\mathfrak{f}}}}\left[\exp\left(2\bigl((1+s)\lambda\bigr)^2\int_0^{T\wedge\zeta_R}\bigl|\hat{J}_s(t)\dot{\boldsymbol{\theta}}(t)\bigr|^2_{\mathbb{R}^d}\,dt\right)\right].\quad\square
\end{aligned}$$

10.5.2. An Expression for Second Covariant Derivatives. As our initial application of the considerations in § 10.5.1, we have the following.

10.69 THEOREM. *Assume that (8.72) holds, let $T\in(0,\infty)$ and $\boldsymbol{\xi}\in\mathbb{R}^d$ be given, and choose $\boldsymbol{\theta}\in C^\infty([0,\infty);\mathbb{R}^d)$ so that $\boldsymbol{\theta}(0)=\boldsymbol{\xi}$ and $\boldsymbol{\theta}(0)=0$. Then,* *for each $R \in (0,\infty)$, there exist an $\epsilon = \epsilon(T, R, \boldsymbol{\theta}) \in (0,\infty)$ such that (cf. (10.67))*

$$\sup_{\substack{s\in[0,1]\\ \mathfrak{f}\in\mathcal{O}(M)}} \mathbb{E}^{\widehat{\mathbb{P}^{\theta}_{\mathfrak{f}}}}\left[\exp\left(\epsilon \sup_{t\in[0,T]}\left|\left[B(\boldsymbol{\theta})\right]_s'(t,\mathbf{W},\mathfrak{P})\right|\right), \zeta_R > T\right] < \infty$$

Hence, for any $f \in C_{\mathrm{b}}(M;\mathbb{R})$ satisfying the growth condition in Corollary 10.51,

$$\begin{aligned}&\mathbb{E}^{\widehat{\mathbb{P}^{\theta}_{\cdot}}}\left[\left(\left[B(\boldsymbol{\theta})\right](T,\mathbf{W}_0,\mathfrak{P}_0)^2 - \left[B(\boldsymbol{\theta})\right]_0'(T,\mathbf{W},\mathfrak{P})\right) f\circ\pi\big(\mathrm{p}(T)\big), \zeta_R > T\right]\\ &\qquad \longrightarrow \mathfrak{E}(\boldsymbol{\xi})^2\big(\mathbf{P}_T^M f\circ\pi\big) \quad \text{as } R\to\infty\end{aligned}$$

in the sense of (L. Schwartz) distributions. In particular, if U is an open subset of M and

$$\sup_{\mathfrak{f}\in\pi^{-1}(U)} \mathbb{E}^{\widehat{\mathbb{P}^{\theta}_{\mathfrak{f}}}}\left[\left|\left[B(\boldsymbol{\theta})\right]_0'(T,\mathbf{W},\mathfrak{P})\right|^2\right] < \infty,$$

then, for $\mathfrak{f}\in\pi^{-1}(U)$,

$$\begin{aligned}&\mathfrak{E}(\boldsymbol{\xi})_{\mathfrak{f}}\circ\mathfrak{E}(\boldsymbol{\xi})\big(\mathbf{P}_T^M f\circ\pi\big)\\ &\qquad = \mathbb{E}^{\widehat{\mathbb{P}^{\theta}_{\mathfrak{f}}}}\left[\left(\left[B(\boldsymbol{\theta})\right](T,\mathbf{W}_0,\mathfrak{P}_0)^2 - \left[B(\boldsymbol{\theta})\right]_0'(T,\mathbf{W},\mathfrak{P})\right) f\circ\pi\big(\mathrm{p}(T)\big)\right].\end{aligned} \tag{10.70}$$

PROOF: The first estimate is an immediate consequence of the estimates in (10.68) and (10.66) combined with (10.60).

To prove the convergence assertion, we start with (10.52) to see that

$$\mathfrak{E}(\boldsymbol{\xi})_{\mathfrak{f}}\circ\mathfrak{E}(\boldsymbol{\xi})\big(\mathbf{P}_T^M f\circ\pi\big) = \frac{d^2}{ds^2}\mathbb{E}^{\widehat{\mathbb{P}^{\theta}_{\mathfrak{f}}}}\left[E_s^{\boldsymbol{\theta}}(T) f\circ\pi\big(\mathfrak{P}_s(T)\big)\right]\Big|_{s=0}. \tag{10.71}$$

Next, note that, by (8.66),

$$\mathbb{E}^{\widehat{\mathbb{P}^{\theta}_{\cdot}}}\left[E_s^{\boldsymbol{\theta}}(T) f\circ\pi\big(\mathfrak{P}_s(T)\big), \zeta_R > T\right] \longrightarrow \mathbf{P}_T^M f\circ\pi$$

uniformly on compacts. At the same time, because $\boldsymbol{\theta}(T) = 0$ and therefore $\pi\circ\mathfrak{P}_s(T) \equiv \pi\circ\mathrm{p}(T)$, the estimates in Lemma 10.63 allow us to justify

$$\begin{aligned}&\frac{d^2}{ds^2}\mathbb{E}^{\widehat{\mathbb{P}^{\theta}_{\mathfrak{f}}}}\left[E_s^{\boldsymbol{\theta}}(T) f\circ\pi\big(\mathfrak{P}_s(T)\big), \zeta_R > T\right]\Big|_{s=0}\\ &\qquad = \mathbb{E}^{\widehat{\mathbb{P}^{\theta}_{\mathfrak{f}}}}\left[\left(\left[B(\boldsymbol{\theta})\right](T,\mathbf{W}_0,\mathfrak{P}_0)^2 - \left[B(\boldsymbol{\theta})\right]_0'(T,\mathbf{W},\mathfrak{P})\right) f\circ\pi\big(\mathrm{p}(T)\big), \zeta_R > T\right]\end{aligned}$$

for each $R\in(0,\infty)$. Thus, the convergence assertion follows from general principles of Schwartz's distribution theory.

Finally, (10.65), (8.66), and the hypothesis stated in the last assertion are more than enough to justify

$$\mathbb{E}^{\widehat{\mathbb{P}^{\boldsymbol\theta}_{\cdot}}}\left[\Big(\big[B(\boldsymbol\theta)\big](T,\mathbf{W}_0,\mathfrak{P}_0)^2-\big[B(\boldsymbol\theta)\big]_0'(T,\mathbf{W},\mathfrak{P})\Big)f\circ\pi(\mathfrak{p}(T)),\ \zeta_R>T\right]$$
$$\longrightarrow\mathbb{E}^{\widehat{\mathbb{P}^{\boldsymbol\theta}_{\cdot}}}\left[\Big(\big[B(\boldsymbol\theta)\big](T,\mathbf{W}_0,\mathfrak{P}_0)^2-\big[B(\boldsymbol\theta)\big]_0'(T,\mathbf{W},\mathfrak{P})\Big)f\circ\pi\big(\mathfrak{p}(T)\big)\right]$$

uniformly on U; and clearly (10.70) follows from this and the preceding. □

10.5.3. Estimates for Derivatives of the Heat Flow. In order to demonstrate how one might use the Theorem 10.69, we will add to (8.72) the assumptions that

$$\tag{10.72}\begin{gathered}\sup_{\mathfrak{f}\in\mathcal{O}(M)}\|\Omega_{\mathfrak f}(\boldsymbol\xi,\boldsymbol\eta)\|_{\mathrm{op}}\le K_\Omega|\boldsymbol\xi|_{\mathbb{R}^d}|\boldsymbol\eta|_{\mathbb{R}^d}\\ \big\|\mathfrak{E}(\boldsymbol\xi)_{\mathfrak f}\mathfrak{R}\big\|_{\mathrm{op}}\le K_{\mathfrak{R}'}\big(1+\mathrm{dist}^M(\pi\mathfrak f,x_0)\big)\end{gathered}$$

for some K_Ω, $K_{\mathfrak{R}'}\in(0,\infty)$ and reference point $x_0\in M$. Under these conditions, it follows from (10.60) and (8.74) that (cf. (10.64))

$$\begin{aligned}&\int_0^T\big|\big(J_0'(t)+A_0(t)J_0(t)\big)\dot{\boldsymbol\theta}(t)\big|^2_{\mathbb{R}^d}\,dt\\&\quad\le T^2\Gamma_\gamma(T,\dot{\boldsymbol\theta})\left[\left(K_{\mathfrak{R}'}\sup_{t\in[0,T]}\frac{\Gamma_\gamma(t,\boldsymbol\theta)}{t}\mathrm{dist}^M\big(\pi\circ\mathfrak{p}(t),x_0\big)\right)^2+\big((d-1)K_\Omega\big)^2\right]\\&\qquad+5\Gamma_\gamma(T,\dot{\boldsymbol\theta})\sup_{t\in[0,T]}\|A_0(t)\|^2_{\mathrm{op}}.\end{aligned}$$

Hence, if we take

$$\boldsymbol\theta(t)=\frac{e^{2\gamma t}-e^{2\gamma T}}{1-e^{2\gamma T}}\boldsymbol\xi,$$

then, from (8.66), (10.66) and (10.68), we see that there exists an $K\in[1,\infty)$, which depends only on d and γ, in addition to K_Ω and $K_{\mathfrak{R}'}$, such that (cf. (6.47))

$$\begin{aligned}&\mathbb{E}^{\widehat{\mathbb{P}^{\boldsymbol\theta}_{\mathfrak f}}}\left[\exp\left(\sup_{t\in[0,T]}\frac{\big|\big[B(\boldsymbol\theta)\big]_0'(t,\mathbf{W},\mathfrak{P})\big|}{K}\right)\right]\\&\qquad\le K\exp\left(\frac{\beta(2\gamma T)^2}{KT}+\big(KT\mathrm{dist}^M(\pi\mathfrak f,x_0)\big)^2\right),\end{aligned}$$

for all $T\in(0,1]$. In particular, (10.70) holds and says that

$$\begin{aligned}&\big|\mathfrak{E}(\boldsymbol\xi)_{\mathfrak f}\circ\mathfrak{E}(\boldsymbol\xi)\big(\mathbf{P}^M_Tf\circ\pi\big)\big|\\&\qquad\le\mathbb{E}^{\widehat{\mathbb{P}^{\boldsymbol\theta}_{\mathfrak f}}}\left[\Big(\big[B(\boldsymbol\theta)\big](T,\mathbf{W}_0,\mathfrak{P}_0)^2+\big|\big[B(\boldsymbol\theta)\big]_0'(T,\mathbf{W},\mathfrak{P})\big|\Big)|f|\circ\pi\big(\mathfrak{p}(T)\big)\right].\end{aligned}$$

Moreover, by the preceding and Lemma 6.45, we know that

$$\mathbb{E}^{\widehat{\mathbb{P}}^{\theta}_{\mathfrak{f}}}\left[\left|[B(\boldsymbol{\theta})]_0'(T,\mathbf{W},\mathfrak{P})\right|\frac{|f|\circ\pi(\mathfrak{p}(T))}{[\mathbf{P}^M|f|](\pi\mathfrak{f})}\right]$$
$$\leq\frac{\beta(2\gamma T)^2}{T}+K^3\big(T\mathrm{dist}^M(\pi\mathfrak{f},x_0)\big)^2+\left[\mathbf{P}^M_T\left(\frac{K|f|}{[\mathbf{P}^M_T|f|](x)}\log\frac{K|f|}{[\mathbf{P}^M_T|f|](x)}\right)\right](x).$$

Similarly, by (10.65) and Lemma 6.45,

$$\mathbb{E}^{\widehat{\mathbb{P}}^{\theta}_{\mathfrak{f}}}\left[[B(\boldsymbol{\theta})](T,\mathbf{W}_0,\mathfrak{P}_0)^2\frac{|f|\circ\pi(\mathfrak{p}(T))}{[\mathbf{P}^M|f|](\pi\mathfrak{f})}\right]$$
$$\leq\frac{\beta(2\gamma T)^2}{T}\left(\left[\mathbf{P}^M_T\left(\log 12+\frac{|f|}{[\mathbf{P}^M_T|f|](x)}\log\frac{|f|}{[\mathbf{P}^M_T|f|](x)}\right)\right](x)\right).$$

Hence, after we combine these two and make a slight readjustment of K, we come to the conclusion contained in the next statement.

10.73 THEOREM. *Assume that (8.72) and (10.72) hold. Then there exists a $K\in(0,\infty)$, which depends only on d, γ, and the constants in (10.72), such that (cf. (6.47),(7.20), and (8.30))*

$$\frac{\|H^M_x\mathbf{P}^M_Tf\|_{\mathrm{op}}}{[\mathbf{P}^M_T|f|](x)}\leq\frac{\beta(2\gamma T)^2}{T}\left(4+\log\frac{\|f\|_{C_b(M;\mathbb{R})}}{[\mathbf{P}^M_T|f|](x)}\right)$$
$$+K\left(\big(\mathrm{dist}^M(x,x_0)\big)^2+\log\frac{K\|f\|_{C_b(M;\mathbb{R})}}{[\mathbf{P}^M_T|f|](x)}\right)\tag{10.74}$$

for all $(T,x)\in(0,1]\times M$. In fact, if the dist^M term can be removed from the right hand side of (10.72), then it can be removed from the right hand side of (10.74).

10.5.4. Estimate on Derivatives of the Heat Kernel. We will now apply (10.74) to the heat kernel (cf. Theorem 6.25) $g_t(x,y)$. Namely, because $g_T(\cdot,y)=\mathbf{P}^M_{\frac{T}{2}}g_{\frac{T}{2}}(\cdot,y)$, we know from (10.74) that

$$\frac{\|H^M_xg_T(\cdot,y)\|_{\mathrm{op}}}{g_T(x,y)}\leq\frac{2\beta(\gamma T)^2}{T}\left(4+\log\frac{\|g_{\frac{T}{2}}(\cdot,y)\|_{C_b(M;\mathbb{R})}}{g_T(x,y)}\right)$$
$$+K\left(\left(\frac{T}{2}T\mathrm{dist}^M(x,x_0)\right)^2+\log\frac{K\|g_{\frac{T}{2}}(\cdot,y)\|_{C_b(M;\mathbb{R})}}{g_T(x,y)}\right)\tag{10.75}$$

for $(T,x)\in(0,2]\times M$.

Clearly, before (10.75) can be considered very satisfactory, it is necessary to estimate the ratio which appears on the right hand side, and this will require us to again invoke results from [27].

10.76 LEMMA. *There exists a* $C \in (0,\infty)$, *which depends only on* d *and* γ^-, *such that*

$$\frac{\|g_T(\,\cdot\,,y)\|_{C_b(M;\mathbb{R})}}{g_{2T}(x,y)} \le C\exp\left(\frac{\operatorname{dist}^M(x,y)^2}{4T}\right), \quad (T,x,y)\in(0,1]\times M^2.$$

PROOF: We begin by writing

$$(*)\qquad \frac{g_T(z,y)}{g_{2T}(x,y)} = \frac{g_T(z,y)}{g_T(y,y)}\frac{g_T(y,y)}{g_{2T}(x,y)}.$$

By Theorem 2.1 in [27] (with the q there identically 0),

$$\frac{g_T(y,y)}{g_{2T}(x,y)} \le C\exp\left(\frac{\operatorname{dist}^M(x,y)^2}{4T}\right), \quad (T,x,y)\in(0,1]\times M^2,$$

for some C with the required dependence. Hence, all that remains is to check that the first factor on the right hand side of (*) is uniformly bounded by a constant depending only on d and γ^-. For this purpose, we first use Corollary 3.1 from [27] to see that (we use $|\Gamma|$ to denote the λ_M-measure of $\Gamma\subseteq M$)

$$g_T(z,y) \le \frac{C}{\sqrt{|B_M(z,T^{\frac12})||B_M(y,T^{\frac12})|}}\exp\left(-\frac{\operatorname{dist}^M(z,y)^2}{4T}\right),$$

for the correct sort of $C\in(0,\infty)$. Second, we use (8.65) to find an $\alpha\in[1,\infty)$, depending only on d and γ^-, so that

$$\mathbb{P}_y^M\left(\operatorname{dist}^M\left(p\left(\tfrac{T}{2}\right),y\right)\ge \alpha T^{\frac12}\right)\le\frac12 \quad \text{for all } (T,y)\in(0,1]\times M.$$

In particular, with this choice of α, we have (by the Chapman–Kolmogorov equation and symmetry) that

$$\begin{aligned} g_T(y,y) &= \int_M g_{\frac{T}{2}}(y,z)g_{\frac{T}{2}}(z,y)\,\lambda_M(dz) \ge \int_{B_M(y,\alpha T^{\frac12})} g_{\frac{T}{2}}(y,z)^2\,\lambda_M(dz) \\ &\ge |B_M(y,\alpha T^{\frac12})|^{-1}\left(\int_{B_M(y,\alpha T^{\frac12})} g_{\frac{T}{2}}(y,z)\,\lambda_M(dz)\right)^2 \ge |4B_M(y,\alpha T^{\frac12})|^{-1}. \end{aligned}$$

where, in the second to last step, we have used Schwarz's inequality. Hence, after combining these two, we arrive at

$$\begin{aligned} \frac{g_T(z,y)}{g_T(y,y)} &\le 4C\frac{|B_M(y,\alpha T^{\frac12})|}{|B_M(y,T^{\frac12})|}\sqrt{\frac{|B_M(y,T^{\frac12})|}{|B_M(z,T^{\frac12})|}}\exp\left(-\frac{\operatorname{dist}^M(z,y)^2}{4T}\right) \\ &\le 4C\frac{|B_M(y,\alpha T^{\frac12})|}{|B_M(y,T^{\frac12})|}\sqrt{\frac{|B_M(z,T^{\frac12}+\operatorname{dist}^M(z,y))|}{|B_M(z,T^{\frac12})|}}\exp\left(-\frac{\operatorname{dist}^M(z,y)^2}{4T}\right). \end{aligned}$$

To complete the proof from here, we use Bishop's Volume Comparison Theorem (cf. Theorem 4.19 in [17]) which says that, for $0 < r \le R$,

$$\frac{|B_M(z,R)|}{|B_M(z,r)|} \le \frac{\int_0^R \left(\sinh\sqrt{K}\rho\right)^{d-1} d\rho}{\int_0^r \left(\sinh\sqrt{K}\rho\right)^{d-1} d\rho} \left(\equiv \left(\frac{R}{r}\right)^d \text{ if } K = 0 \right),$$

where $K \in [0,\infty)$ is determined so that $\gamma^- = (d-1)K$. Since, by an elementary calculation, this leads to

$$\frac{|B_M(z,R)|}{|B_M(z,r)|} \le \cosh\sqrt{K}r \left(\frac{\sinh\sqrt{K}R}{\sinh\sqrt{K}r}\right)^d \le \left(\cosh\sqrt{K}r\right)\exp\left(d\sqrt{K}\frac{R}{r}\right)$$

when $K > 0$ and $R > r \in (0,1]$, the required bound follows easily. □

10.77 THEOREM. *Assume that (8.72) holds. Then there is a $C \in (0,\infty)$, depending only on d and γ^-, such that (cf. (6.47))*

$$\frac{\left\|\operatorname{grad}_x^M g_T(\,\cdot\,,y)\right\|}{g_T(x,y)} \le \frac{2\beta(\gamma T)\operatorname{dist}^M(x,y)}{T} + C$$

for all $(T,x,y) \in (0,1] \times M^2$. If, in addition, (10.72) holds, then there exists a $K \in (0,\infty)$, depending only on d, γ, and the constants in (10.72), such that

$$\begin{aligned}\frac{\|H_x^M g_T(\,\cdot\,,y)\|_{\mathrm{op}}}{g_T(x,y)} \le & 2\left(\frac{\beta(\gamma T)\operatorname{dist}^M(x,y)}{T}\right)^2 \\ & + K\left(\frac{1+\operatorname{dist}^M(x,y)^2}{T} + T\operatorname{dist}^M(x,x_0)^2\right)\end{aligned}$$

for all $(T,x,y) \in (0,1] \times M^2$. In particular, if the term $\operatorname{dist}^M(\pi\mathfrak{f},x_0)$ can be removed from the right hand side of the second line in (10.72), then the terms involving $\operatorname{dist}^M(x,x_0)$ can be removed from the right hand side of the preceding.

PROOF: As we pointed out at the close of § 8.5.3, once we have Bismut's formula (8.80) and the estimate in (8.79), the estimate in Theorem 6.46 is an immediate corollary. In particular, by taking $f = g_{\frac{T}{2}}(\,\cdot\,,y)$, we know that

$$\frac{\left\|\operatorname{grad}_x^M g_T(\,\cdot\,,y)\right\|}{g_T(x,y)} \le \frac{2^{\frac{3}{2}}\beta(\gamma T)}{T^{\frac{1}{2}}}\sqrt{\log\frac{\|g_{\frac{T}{2}}(\,\cdot\,,y)\|_{C_b(M;\mathbb{R})}}{g_T(x,y)}}.$$

Hence, the first part of the present theorem follows immediately upon plugging in the estimate from Lemma 10.76. Similarly, the second estimate is an immediate consequence of the estimates in Lemma 10.76 and (10.75). □

Remark: It is gratifying that the estimates in Theorem 10.77 are quite sharp in the sense that one cannot do much better even in the case when $M = \mathbb{R}^d$. Similar estimates on higher derivatives can be found in [40]. In this connection, it should be recognized that more interesting phenomena are encountered when one studies

$$H_x^M \log g_T(\,\cdot\,,y) = \frac{H_x^M g_T(\,\cdot\,,y)}{g_T(x,y)} - \frac{\operatorname{grad}_x^M g_T(\,\cdot\,,y) \otimes \operatorname{grad}_x^M g_T(\,\cdot\,,y)}{g_T(x,y)^2}.$$

Indeed, a careful analysis shows that, as long as x stays away from the cut-locus of y, there is sufficient cancellation between the preceding terms to kill the T^{-2} which appears in the second part of Theorem 10.77. On the other hand, when x is in the cut-locus of y, this cancellation will not take place and the behavior of $H_x^M g_T(\,\cdot\,,y)$ can be as bad as T^{-2}. For example, when M is the standard unit circle in the plane, then explicit calculation shows that

$$H_x^M \log g_T(\,\cdot\,,y) \sim \begin{cases} \frac{\pi^2}{T^2} & \text{if } x \text{ the polar opposite of } y \\ -\frac{1}{T} & \text{otherwise.} \end{cases}$$

A detailed analysis (based on the sort of reasoning alluded to at the end of § 1.3.1) of this sort of phenomenon is given in [29], and the case of higher derivatives is studied in [41].

10.6 An Admission of Defeat

One would hope that the machinations in the preceding section might be sufficient to allow one to recover Li and Yau's basic differential inequality (cf. Theorem 1.2 of [27] and Lemma 6.26 here) which says that

$$(10.78)\qquad \begin{gathered} \frac{\|\operatorname{grad}_x \mathrm{P}_T^M f\|^2}{(\mathrm{P}_T^M f)^2} - \alpha\frac{\Delta_M \mathrm{P}_T^M f}{\mathrm{P}_T^M f} \le \frac{d\alpha^2}{T} + \frac{2^{\frac{1}{2}} d\alpha^2 \gamma^-}{\alpha - 1} \\ \text{for } f \in C\big(M;(0,\infty)\big) \text{ and } \alpha \in (1,2), \end{gathered}$$

where one can and should take $\alpha = 1$ when $\gamma \ge 0$. However, in spite of wasting considerable time and effort in an attempt to do so, I have failed. The problem comes from the term $[B(\theta)]_0^\prime(T, \mathbf{W}, \mathfrak{P})$ on the right hand side of (10.71). Indeed, as we have already seen, this quantity contains terms whose control requires more than estimates on the lower bound of the Ricci curvature, and yet the lower bound on the Ricci curvature is all that is needed for the Li–Yau estimate. Of course, (10.71) is an expression for an arbitrary, individual, second, covariant derivative, whereas the Li–Yau deals only with the Laplacian. Thus, there is a possibility that some miraculous cancellation occurs when one takes the trace of the Hessian, but, if it exists at all, then that cancellation has successfully eluded me except in the special case explained below.

10.6.1. Li and Yau for Einstein Manifolds. The only case in which I have been able to prove (10.78) is when M is an *Einstein manifold* with non-positive Ricci curvature. That is, I have to assume that

$$\mathfrak{R} \equiv 2\gamma I_{\mathbb{R}^d} \quad \text{with } \gamma \in (-\infty, 0]. \tag{10.79}$$

It should be obvious why the Einstein condition greatly simplifies our lives. Namely, it means that (cf. (8.73))

$$\mathfrak{J}_{\mathfrak{p}}(t) = e^{-\gamma t} I_{\mathbb{R}^d}. \tag{10.80}$$

Hence, if we assume that θ has the form $\theta(t) = \theta(t)\xi$, then

$$\big[B(\theta)\big](T, \mathbf{W}_s, \mathfrak{P}_s) = \int_0^T e^{-\gamma t}\dot{\theta}(t)\, d\big(\xi, \mathbf{W}_s(t)\big)_{\mathbb{R}^d}.$$

In particular, by taking $\theta_k(t) = \theta(t)\mathrm{e}_k$ where $\theta(0) = 1$ and $\theta(T) = 0$, we have, from (10.55) and Theorem 10.69, that

$$\frac{\|\mathrm{grad}\mathbf{P}_T^M f\|^2 \circ \pi}{[\mathbf{P}_T^M f]^2 \circ \pi} - \alpha \frac{\big[\Delta_M \mathbf{P}_T^M f\big] \circ \pi}{\mathbf{P}_T^M f} \le \varlimsup_{R\to\infty} F_R(T, \cdot) \tag{10.81}$$

when

$$\begin{aligned} F_R(T, \mathfrak{f}) \equiv &-\frac{\alpha - 1}{[\mathbf{P}_T^M f] \circ \pi} \mathbb{E}^{\widehat{\mathbb{P}^{\theta_k}_{\cdot}}} \left[\left| \int_0^T e^{-\gamma t}\dot{\theta}(t)\, d\mathbf{W}_0(t) \right|^2_{\mathbb{R}^d} f \circ \pi\big(\mathfrak{p}(T)\big),\ \zeta_R > T \right] \\ &+ \frac{1}{\mathbf{P}_T^M f(x)} \sum_{k=1}^d \mathbb{E}^{\widehat{\mathbb{P}^{\theta_k}_{\mathfrak{f}}}} \left[\big[B(\theta_k)\big]_0'(T, \mathbf{W}, \mathfrak{P}) f \circ \pi\big(\mathfrak{p}(T)\big),\ \zeta_R > T \right] \end{aligned}$$

where the inequality results from our having thrown away the sum over $1 \le k \le d$ of the quantities

$$\left(\frac{\mathbb{E}^{\widehat{\mathbb{P}^{\theta_k}_{\mathfrak{f}}}} \big[[B(\theta_k)](T, \mathbf{W}_0, \mathfrak{P}_0) f \circ \pi\big(\mathfrak{P}_0(T)\big),\ \zeta_R > T \big]}{\mathbf{P}_T^M f(\pi \mathfrak{f})} \right)^2 - \frac{\mathbb{E}^{\widehat{\mathbb{P}^{\theta_k}_{\mathfrak{f}}}} \big[[B(\theta_k)](T, \mathbf{W}_0, \mathfrak{P}_0)^2 f \circ \pi\big(\mathfrak{P}_0(T)\big),\ \zeta_R > T \big]}{\mathbf{P}_T^M f(\pi \mathfrak{f})},$$

each of which, by Schwarz's inequality, is non-positive.

In order to take the next step, we will assume, at least for the moment, that all the paths are piecewise smooth, in which case (because of (10.80)), we can write

$$\big[B(\theta_k)\big]_0'(T, \mathbf{W}, \mathfrak{P}) = \int_0^T e^{-\gamma t}\dot{\theta}(t)\big(\mathrm{e}_k, \dot{\mathbf{W}}_0'(t)\big)_{\mathbb{R}^d}\, dt \quad \widehat{\mathbb{P}^{\theta_k}_{\mathfrak{f}}}\text{-almost surely.}$$

At the same time, by Cartan's first structural equation (8.48),

$$\dot{\mathbf{W}}_0'(t) = \frac{d}{ds}\phi(\dot{\mathfrak{P}}_s(t))\Big|_{s=0} = \frac{d}{dt}\phi(\mathfrak{P}_0'(t)) - \omega(\mathfrak{P}_0'(t))\dot{\mathbf{W}}_0(t),$$

and, $\widehat{\mathbb{P}_{\mathfrak{f}}^{\theta_k}}$-almost surely,

$$\frac{d}{dt}\phi(\mathfrak{P}_0'(t)) = \frac{d}{dt}\big(e^{-\gamma t}\theta(t)\big)\mathbf{e}_k,$$

while, with the help of Cartan's second structural equation (8.49),

$$-\omega(\mathfrak{P}_0'(t))\dot{\mathbf{W}}_0(t) = \int_0^t e^{-\gamma\tau}\theta(\tau)\Omega_{\mathfrak{P}_0(\tau)}\big(\mathbf{e}_k, \dot{\mathbf{W}}_0(\tau)\big)\dot{\mathbf{W}}_0(t)\,d\tau.$$

Thus, because of (8.58) and (10.79), we are predicting that

$$\begin{aligned}
&\varlimsup_{R\to\infty} \frac{1}{\mathbf{P}_T^M f(x)} \sum_{k=1}^{d} \mathbb{E}^{\widehat{\mathbb{P}_{\mathfrak{f}}^{\theta_k}}} \Big[\big[B(\theta_k)\big]_0'(T, \mathbf{W}, \mathfrak{P}) f\circ\pi(\mathfrak{P}_0(T)),\, \zeta_R > T\Big] \\
&= d\int_0^T e^{-2\gamma t}\big(\dot\theta(t) - \gamma\theta(t)\big)\dot\theta(t)\,dt \\
&\quad + 2\gamma\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\iint\limits_{0\le\tau\le t\le T} \theta(\tau)\dot\theta(t)\big(e^{-\gamma\tau}\dot{\mathbf{w}}(\tau), e^{-\gamma t}\dot{\mathbf{w}}(t)\big)_{\mathbb{R}^d}\,d\tau dt\,\frac{f\circ\pi(\mathfrak{p}(T,\mathfrak{f},\mathbf{w}))}{\mathbf{P}_T^M f(x)}\right].
\end{aligned}$$

Finally, if we write

$$\begin{aligned}
&2\gamma \iint\limits_{0\le\tau\le t\le T} \theta(\tau)\dot\theta(t)\big(e^{-\gamma\tau}\dot{\mathbf{w}}(\tau), e^{-\gamma t}\dot{\mathbf{w}}(t)\big)_{\mathbb{R}^d}\,d\tau dt \\
&\quad = -\gamma\int_0^T \frac{\theta(\tau)}{\dot\theta(\tau)}\frac{d}{d\tau}\left|\int_\tau^T \dot\theta(t)e^{-\gamma t}\dot{\mathbf{w}}(t)\,dt\right|_{\mathbb{R}^d}^2 d\tau \\
&\quad = \frac{\gamma}{\dot\theta(0)}\left|\int_0^T \dot\theta(t)e^{-\gamma t}\dot{\mathbf{w}}(t)\,dt\right|_{\mathbb{R}^d}^2 \\
&\qquad\qquad + \gamma\int_0^T \frac{d}{d\tau}\left(\frac{\theta(\tau)}{\dot\theta(\tau)}\right)\left|\int_\tau^T \dot\theta(t)e^{-\gamma t}\dot{\mathbf{w}}(t)\,dt\right|_{\mathbb{R}^d}^2 d\tau,
\end{aligned}$$

then, after we plug these into (10.81), what we are predicting is that

$$
\begin{aligned}
&\frac{\|\operatorname{grad}_x \mathrm{P}_T^M f\|^2}{(\mathrm{P}_T^M f)^2} - \alpha \frac{\Delta_M \mathrm{P}_T^M f}{\mathrm{P}_T^M f} \\
&\quad \le d \int_0^T e^{-2\gamma t}\big(\dot\theta(t) - \gamma\theta(t)\big)\dot\theta(t)\,dt \\
&\quad + \left(\frac{\gamma}{\dot\theta(0)} + 1 - \alpha\right)\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\left|\int_0^T e^{-\gamma t}\dot\theta(t)\,d\mathbf{w}(t)\right|_{\mathbb{R}^d}^2 \frac{f\circ\pi\big(\mathfrak{p}(T,\mathfrak{f},\mathbf{w})\big)}{\mathrm{P}_T^M f(x)}\right] \\
&\quad + \gamma\mathbb{E}^{\mu_{\mathbb{R}^d}}\left[\int_0^T \frac{d}{d\tau}\left(\frac{\theta(\tau)}{\dot\theta(\tau)}\right)\left|\int_\tau^T e^{-\gamma t}\dot\theta(t)\,d\mathbf{w}(t)\right|_{\mathbb{R}^d}^2 d\tau \frac{f\circ\pi\big(\mathfrak{p}(T,\mathfrak{f},\mathbf{w})\big)}{\mathrm{P}_T^M f(x)}\right].
\end{aligned}
\tag{10.82}
$$

Of course, this prediction was made under the unacceptable assumption that our paths are smooth. On the other hand, because the right hand side of (10.82) makes perfectly good sense for any $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$, we can apply our usual polygonalization procedure to justify the all steps taken.

Given (10.82), the rest is easy. Namely, we take

$$\theta(t) = \left(1 - \frac{t}{T}\right)e^{\gamma t},$$

Then (remember that $\gamma \le 0$):

$$\int_0^T e^{-2\gamma t}\big(\dot\theta(t) - \gamma\theta(t)\big)\dot\theta(t)\,dt = \frac{1}{T} + \frac{|\gamma|}{2},$$

$$\frac{\gamma}{\dot\theta(0)} + 1 - \alpha = \frac{(2-\alpha)|\gamma|T - (\alpha - 1)}{1 + |\gamma|T} \le 0 \quad \text{when } T \le T_\gamma(\alpha) \equiv \frac{\alpha - 1}{(2-\alpha)|\gamma|},$$

and

$$\frac{d}{d\tau}\left(\frac{\theta(\tau)}{\dot\theta(\tau)}\right) = \big(1 + |\gamma|(T - \tau)\big)^{-2} \ge 0 \quad \text{for } 0 \le \tau \le T.$$

Hence, by (10.82), we see that the left hand side of (10.78) is dominated by $\frac{d}{T} + \frac{d|\gamma|}{2}$ as long as $T \le T_\gamma(\alpha)$. Since we can write $\mathrm{P}_T^M = \mathrm{P}_{T_\gamma(\alpha)}^M \circ \mathrm{P}_{T - T_\gamma(\alpha)}^M$ when $T > T_\gamma(\alpha)$, we can now say that (10.79) implies

$$\frac{\|\operatorname{grad}_x \mathrm{P}_T^M f\|^2}{(\mathrm{P}_T^M f)^2} - \alpha \frac{\Delta_M \mathrm{P}_T^M f}{\mathrm{P}_T^M f} \le \frac{d}{T} + \frac{d(3-\alpha)|\gamma|}{2(\alpha - 1)}. \tag{10.83}$$

On the one hand, (10.83) is a slight improvement on (10.78). On the other hand, we have proved it only under very rigid conditions.

1. Bakry, D., *L'hypercontractivité et son utilisation en théorie des semigroupes*, Lectures on probability theory (Saint-Flour, 1992), LNM #1581, Springer–Verlag, 1994, pp. 1–114.
2. Bishop, R. & Crittenden, R., *Geometry of Manifolds*, Pure & Appl. Math. Series #15, Academic Press, 1964.
3. Bismut, J.–M., *Large Deviations and the Malliavin Calculus*, Progress in Math. #45, Birkhäuser, 1984.
4. Calabi, E., *An extension of E. Hopf's maximum principle with an application to Riemannian geometry*, Duke Math. J. **25** (1957), 45–56.
5. Chavel, I., *Riemannian Geometry: a Modern Introduction*, Cambridge Tracts in Math. #108, Cambridge U. Press, 1993.
6. Cranston, M., *Gradient estimates on manifolds using coupling*, J. Fnal. Anal. **99** (1991), 110–124.
7. Deuschel, J-D & Stroock, D., *Large Deviations*, Pure & Appl. Math. Series **#137**, Academic Press, 1989.
8. do Carmo, M., *Riemannain Geometry*, Mathematics: Theory & Applications, Birkhäuser, 1993.
9. Driver, B., *A Cameron-Martin type quasi-invariance theorem for Brownian motion on compact Riemannian manifolds*, Jour. Funct. Anal. **110** (1992), 272–376.
10. Echverria, P., *A criterion for invariant measures of a Markov process*, Zeit. Wahr. und Verw. Gebiete **62#1** (1982), 1–16.
11. Eells, J. & Elworthy, D., *Stochastic dynamical systems*, Control Theory and Topics in Functional Analysis, Vol. III, Intern. Atomic Energy Agency, Vienna, 1979, pp. 179–185.
12. Einstein, A., *On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat* [English tranlation of German title], English translation in *Investigations on the theory of the Brownian movement*, Edited by R. Fürth and published by Dutton in 1926, Ann. d. Physik (ser. 4) **17** (1905), 549–560.
13. Elworthy, K.D. & Yor, M., *Conditional expectations for the derivatives of certain stochastic flows*, Sémaire de Probabilités XXVII, LNM 1557, Springer-Verlag, 1993, pp. 159–172.
14. Emery, M., *Stochastic Calculus in Manifolds*, Universitext Series, Springer–Verlag, 1989.
15. Enchev, O. and Stroock, D., *Towards a Riemannian geometry on the path space over a Riemannian manifold*, J. Fnal. Anal. **134** (1995), 392–416.
16. Friedman, A., *Partial Differential Equations of the Parabolic Type*, Prentice Hall, 1964.
17. Gallot, S., Hulin, D., and Lafontaine, J., *Riemann Geometry, 2nd. ed.*, Universitext Series, Springer–Verlag, 1990.
18. Gangolli, R., *On the construction of certain diffusions on a differentiable manifold*, Zeit. Wahr. und Verw. Gebiete **2** (1964), 406–419.
19. Grigor'yan, A.A., *On stochastically complete manifolds*, Soviet Math. Dokl. (English translation) **34 #2** (1987), 310–313.
20. ———, *Analytic and geometric background of recurrent and non-explosion of the Brownian motion on Riemannian manifolds*, BAMS **32 #2** (1999), 135–251.
21. Hörmander, L., *Hypoelliptic second order differential operators*, Acta Math. **121** (1968), 147–171.
22. Ikeda, N. & Watanabe, S., *Stochastic Differential Equations and Diffusion Processes*, North-Holland/Kodansha, 1981.
23. Kendell, W., *Non-negative Ricci curvature and the Brownian coupling property*, Stochastics **19** (1986), 110–124.
24. Kobayashi, S. & Nomizu, K., *Foundations of Differential Geometry, I & II*, Iterscience Tracts in Pure & Appl. Math. #15, J. Wiley, 1969.
25. Ladyženskaja, O.A., Solonnikov, V.A., & Uralćeva, *Linear and Quasilinear Equations of Parabolic Type*, Translations of Math. Monographs #23, AMS, 1968.

26. Lewis, J., *Brownian motion on a submanifold of Euclidean Space*, Bull. London Math. Soc. (1984), 144–150.
27. Li, P. & Yau, S.-T., *On the parabolic kernel of the Schrödinger operator*, Acta. Math. **156** (1986), 153–201.
28. Malliavin, P., *Formule de la moyenne, calcul des perturbations, et théorème d'annulation pour les formes harmoniques*, J. Fnal. Anal. **17** (1974), 274–291.
29. Malliavin, P. & Stroock, D., *Short time behavior of the heat kernel and its logarithmic derivatives*, J. Diff. Geom. **44** (1996), 550–570.
30. Milnor, J., *Morse Theory*, Annals of Math. Studies #51, Princton U. Press, 1963.
31. Nash, J., *The embedding theorem for Riemannian manifolds*, Ann. of Math. **63** (1956), 20–63.
32. Pinsky, M., *Brownian motion and Riemannian geometry*, Probability Theory (Chen, L, Choi, K.P., Hu, K., & Lou, J.-H., eds.), de Gruyter, 1991.
33. Rogers, L.C.G. & Williams, D,, *Diffusions, Markov Processes, and Martingales, vol. 2*, J. Wiley, 1987.
34. Spivak, M., *A Comprehensive Introduction to Riemannian Geometry, II*, Publish or Perish, Inc., 1970.
35. Stroock, D., *Probability Theory, An Analytic View*, Cambridge Univ. Press, 1993.
36. ______, *A Concise Introductian to the Theory of Integratian*, Birkhäuser, 1994.
37. ______, *Non-divergence form operators and variations on Yau's explosion criterion*, J. Fourier Anal. & Appl. **4** #**4** & **5** (1998), 565–574.
38. ______, *On the growth of stochastic integrals*, Zeit. Wahr. und Verw. Gebiete **18** (1971), 340–344.
39. Stroock, D. & Taniguchi, S., *Diffusions as integral curves, or Stratonovich without Itô* (1994), Birkhäuser, 331–369.
40. Stroock, D. & Turetsky, J., *Upper bounds on derivatives of the logarithm of the heat kernel*, Comm. Anal. & Geom. 6 #**4** (1998), 669–685.
41. ______, *Short time behavior of logarithmic derivatives of the heat kernel*, Asian J. Math. **1** #**1** (1997), 17–33.
42. Stroock, D. & Varadhan, S.R.S., *Multidimensional Diffusion Processes*, Grundlehren Series #233, Springer–Verlag, 1979 & 1998.
43. Taylor, M., *Pseudodifferential Operators*, Princeton Math. Series #34, Princeton U. Press, 1981.
44. Whitney, H., *Differentiable manifolds*, Ann. of Math. **37** (1936), 645–680.
45. Wiener, N., *Differential space*, J. Math. & Physics **2** (1923), 131–174.
46. Yau, S.-T., *On the heat kernel of a complete Riemannian manifold*, J. Math. Pures Appl., ser. 9 **57** (1978), 191–201.

Index

F

G

H

I

J

K

L

M

N

O

P

R

S

T

V

W