Stochastic Representation of the Gradient and Hessian of Diffusion Semigroups on Riemannian Manifolds

Dissertation zur Erlangung des Doktorgrades
der Naturwissenschaften (Dr. rer. nat.) der
Naturwissenschaftlichen Fakultät I – Mathematik
der Universität Regensburg

vorgelegt von
Holger Plank
aus Kelheim

Regensburg, im Juli 2002

Promotionsgesuch eingereicht am:		10. Juli 2002
Diese Arbeit wurde angeleitet von:		Prof. Dr. Anton Thalmaier
Prüfungsausschuss:	Vorsitzender:	Prof. Dr. Knut Knorr
	1. Gutachter:	Prof. Dr. Anton Thalmaier
	2. Gutachter:	Prof. Dr. Wolfgang Hackenbroch
	Weiterer Prüfer:	Prof. Dr. Felix Finster Zirker
Termin der mündlichen Prüfung:		20. Dezember 2002

Contents

Introduction

One of the best known examples of direct relations between analytic and probabilistic objects is the theory of second order elliptic differential operators (in particular, the Laplacian) on the analytic side and the theory of *diffusion processes* for the stochastic part (with Brownian motion being the most famous representative).
If one uses the semimartingale access to a (for simplicity) $\mathbb{R}^n$-valued diffusion X_t solving a stochastic differential equation

$$\mathrm{d}X_t = \sigma(X_t)\mathrm{d}B_t + \beta(X_t)\mathrm{d}t \tag{1}$$

with an $\mathbb{R}^m$-valued Brownian motion B_t as well as smooth maps $x \mapsto \sigma(x) \in \mathrm{Lin}(\mathbb{R}^m, \mathbb{R}^n)$ and $x \mapsto \beta(x) \in \mathbb{R}^n$ on $\mathbb{R}^n$, then by means of Itô's stochastic calculus, for any real function $f \in C_c^\infty(\mathbb{R}^n)$, the probabilistic average $\mathbb{E}\left[f(X_t)\right]$ is related to a second order partial differential operator on $\mathbb{R}^n$ the following way:
Assumed that the diffusion starts almost surely in a fixed point $X_0 = x$, we have

$$\mathbb{E}[f(X_t)] - f(x) = \mathbb{E}\left[\int_0^t Lf(X_s)\mathrm{d}s\right], \tag{2}$$

where

$$L = \sum_{i,j=1}^n (\sigma\sigma^*)_{ij} \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^n \beta_i \frac{\partial}{\partial x_i}. \tag{3}$$

Obviously the right hand side of (2) vanishes for an L-harmonic function f, and since (2) generalizes to $f \in C^2(\mathbb{R}^n)$ when t is replaced by a suitable random time (e.g. the first exit time of X from a ball around x), there are immediate applications like *boundary integral representations of L-harmonic functions*.
Furthermore, if instead of f one allows time-dependent functions $F : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}$, which adds an additional *drift term* involving $\frac{\partial}{\partial t}F$ on the right hand side of (2), the same argument leads to a stochastic representation of space-time-harmonic functions (i.e. $\left(\frac{\partial}{\partial t} + L\right)F = 0$) on $\mathbb{R}^n$. Of course, this particularly applies to the *diffusion* or *heat semigroup* case $F_t = P_{T-t}f$ for some fixed time horizon T; P_{T-t} denoting the (minimal) heat semigroup on $\mathbb{R}^n$ with respect to L, here acting on $L^\infty(\mathbb{R}^n)$.
Then the stochastic representation of $P_t f$ (in the non-explosive situation) is given by

$$P_t f(x) = \mathbb{E}[f \circ X_t(x)], \tag{4}$$

where $X_t(x)$ is the solution of (1) with $X_0(x) = x$.

The present work is based on extensions of Itô's diffusion theory. On the one hand side, there exists the rich topic of *stochastic analysis on Riemannian manifolds*, and on the other hand, there are many results about parameter-dependance of families of semimartingales, in particular for *stochastic flows*. For instance, we make substantial use of the fact that our diffusion processes $X_t(x)$ depend smoothly on their (deterministic) initial value x.

In 1984, Bismut proved the following remarkable fact ([Bi], Theorem 2.14): On a compact and connected Riemannian manifold M, one has a stochastic representation of the quotient $\frac{\mathrm{grad}_x\, p(t,\,\cdot\,,y)}{p(t,x,y)}$ for the (smooth) heat kernel $p(t,x,y)$ with respect to the Laplace-Beltrami operator on M, which reads

$$\frac{\mathrm{grad}_x\, p(t,\,\cdot\,,y)}{p(t,x,y)} = \frac{1}{t}\mathbb{E}^{\mathbb{P}^t_{x,y}}\left[\int_0^t \widetilde{E}'_s \mathrm{d}\beta_s\right]. \tag{5}$$

Herein $\mathbb{P}^t_{x,y}$ denotes the *Brownian bridge measure* obtained by conditioning Brownian paths on M to start in x and run into y within time t. β denotes a Brownian motion and $\widetilde{E}'$ a certain semimartingale, both taking values in the tangent space T_xM.
The possibility to express the gradient of the logarithmic heat kernel by the expectation of a (conditioned) stochastic integral aroused the interest of quite a lot of stochastic analysts and thus at the present time there are many related results which are often referred to as "Bismut (type) formulae".

The topic was picked up again by Elworthy in [El 3], who chose a rather elementary way to prove heat semigroup derivative formulae instead of using Malliavin calculus as Bismut did. A more systematical treatment by Elworthy and Li appeared in [E-L 1].
The last mentioned article starts out by "formulae with simple proof for $\mathbb{R}^n$" which gives an occasion to illustrate their basic idea.

Consider the *Stratonovich stochastic differential equation* (the Stratonovich context is preferable for transferring results to the geometrical situation afterwards) on $\mathbb{R}^n$

$$\mathrm{d}X_t = A(X_t) * \mathrm{d}B_t + A_0(X_t)\mathrm{d}t, \tag{6}$$

where A and A_0 fulfill the same conditions as the coefficients σ and β did in (1). For simplicity we assume that the (pointwise) adjoint $A^*(x)$ is a right inverse of $A(x)$, which causes the generator of X_t to equal

$$L = \frac{1}{2}\Delta + V$$

for some first order partial differential operator V. Further assume that solutions to (6) are globally defined.
Now for $f : \mathbb{R}^n \to \mathbb{R}$ being bC^1 (bounded and C^1 with bounded gradient) the space-time-harmonicity of $P_{T-t}f$ implies

$$f(X_T(x)) = P_T f(x) + \int_0^T d(P_{T-s}f)_{X_s(x)} A(X_s(x))\mathrm{d}B_s, \tag{7}$$

where $X_s(x)$ denotes the solution to (6) with initial value $X_0(x) = x$ and $d(P_sf)_y$ the differential of the smooth function P_sf at $y \in \mathbb{R}^n$.
We denote by v_t the *derivative process* to $X_t(x)$ which means that $v_t = (dX_t)_x v_0$ for some given $v_0 \in \mathbb{R}^n$ (it can be shown that v_t is – up to modification – given as solution to the stochastic differential equation derived by the formal differentiation of (6)).

Moreover, we get a heat semigroup $P_t^{(1)}$ on 1-forms by letting

$$P_t^{(1)}(\varphi) : \mathbb{R}^n \to \mathrm{Lin}(\mathbb{R}^n, \mathbb{R}), \qquad (P_t^{(1)}(\varphi))_x(v_0) := \mathbb{E}[(\varphi)_{X_t(x)}(v_t)], \tag{8}$$

$\varphi \in \mathrm{Lin}(\mathbb{R}^n, \mathbb{R})$. If now the formal differentiation under the expectation (according to the stochastic representation (4) of the heat semigroup)

$$d(P_t f)_x(v_0) = (P_t^{(1)}(df))_x(v_0) \tag{9}$$

holds true and $\int_0^t \langle A^*(X_s(x))v_s, \mathrm{d}B_s\rangle$ is a martingale for times $t \in [0, T]$, then the product rule applied to (7) yields

$$\begin{aligned}
\mathbb{E}\left[f(X_T(x)) \int_0^T \langle A^*(X_s(x))v_s, \mathrm{d}B_s\rangle\right] &= \mathbb{E}\left[\int_0^T d(P_{T-s}f)_{X_s(x)} v_s \mathrm{d}s\right] \\
&= \mathbb{E}\left[\int_0^T (P_{T-s}^{(1)}(df))_{X_s(x)}(v_s)\mathrm{d}s\right] \\
&= \int_0^T \left((P_s^{(1)})(P_{T-s}^{(1)}(df))_x\right)(v_0)\mathrm{d}s \\
&= \int_0^T (P_T^{(1)}(df))_x(v_0)\mathrm{d}s = T P_T^{(1)}(df)_x(v_0)
\end{aligned}$$

by the aid of (8) for $\varphi = P_{T-s}^{(1)}(df)$ and the semigroup property of $P_s^{(1)}$. So we finally arrive at the stochastic representation formula

$$\langle \mathrm{grad}_x P_t f, v_0\rangle = d(P_t f)_x(v_0) = \frac{1}{t}\mathbb{E}\left[f(X_t(x)) \int_0^t \langle A^*(X_s(x))v_s, \mathrm{d}B_s\rangle\right]. \tag{10}$$

Formulae of this type as well as a similar result for the *Hessian*, both transferred to the manifold-valued case, can be found [E-L 1]. It should be mentioned that in this context a result comparable to (5) can be derived as a corollary to the representation formula (10).

In the following, we assume X to take values in a finite dimensional Riemannian manifold M.

The procedure from above is not satisfying for several reasons. First of all, note that (10) does not involve derivatives of f, which is in accordance with elliptic regularity ensuring that $P_t f \in C^\infty(M)$ for $t > 0$ even if f is only bounded and measurable. Thus there should be a proof of (10) not related to any differentiability of f. Of course, (9) does not make sense for f not being differentiable. Moreover, if the diffusion is explosive (e.g. if one wants to treat non-compact manifolds), we have

$$P_t f(x) = \mathbb{E}[f \circ X_t(x) 1_{t<\zeta(x)}],$$

$\zeta(x)$ the lifetime of $X_t(x)$, and then (9) may fail for smooth f. Considerations of that type are found in [Th 1] at the end of the first section. For an explicit example see [Th 2]. Another question is how to include situations with boundary into (10), which for instance already occur if one has to stop $X_t(x)$ when exiting a certain domain. Taking $\frac{1}{t}$ under the expectation and replacing t by the stopping time is not adequate because this spoils the martingale property of the stochastic integral.

These problems can be overcome by generalizing v_0/t in the process $v_s/t \equiv T_xX_sv_0/t$ (with T_xX_s the derivative process of $X_s(x)$) to the time derivative of a suitable *finite energy process* K_s taking values in T_xM. Such a modified version of the stochastic representation formula proven by disturbing the underlying diffusion and compensating the resulting drift by a Girsanov change of measure was given by Thalmaier in [Th 1], where the author as well provided an explicit construction of a suitable finite energy process (which is a non-trivial problem in the domain case).
Another improvement, already mentioned in [E-L 1], is to use the *damped*, or also called *Dohrn-Guerra transport* W_s along the paths of $X_s(x)$ instead of T_xX_s which makes the Bismut formulae *intrinsic* in the following sense. The Stratonovich equation (6) may carry "redundant noise", which means that if the dimension m of the driving Brownian motion B_t is greater than the dimension n of the manifold (in which the diffusion takes its values), the filtration generated by B_t is larger than the one generated by $X_t(x)$ itself. In the presence of redundant noise due to a non-trivial kernel of A, the right hand side of the representation formula depends on the choice of the coefficients A and A_0 of the equation and not only on the Riemannian metric on M and related objects like the generator L and curvature terms. (Note that T_xX_s depends on A and not only on the generator of the diffusion.)
In this situation the use of W_s corresponds to the so-called procedure of "filtering out redundant noise", or, more exactly: W_s is the conditional expectation of T_xX_s with respect to the filtration generated by $X_\bullet(x)$, given that either one uses the so-called *Le Jan-Watanabe connection* ∇^{LJW} on TM to define the damped transport, or one assumes that equation (6) describes a *gradient Brownian system*, where ∇^{LJW} equals the Levi-Civita connection on the Riemannian manifold.
The facts about filtering noise were studied in [E-Y], and for a recent treatment of the Le Jan-Watanabe connection we refer to [E-LJ-L].

As direct applications of the stochastic representation formulae for the gradient (depending on a finite energy process), in [Th-W] Thalmaier and Wang obtained *gradient estimates* of the form

$$|\operatorname{grad} u(x)| \leq C\|u\|_D$$

for $u : D \to \mathbb{R}_+$ being L-harmonic on a regular open domain $x \in D \subset M$, which in the positive case $u \geq 0$ can be modified to

$$|\operatorname{grad} u(x)| \leq C\sqrt{u(x)\ \|u\|_D}.$$

The constant C merely depends on a lower *Ricci curvature* bound on D, the dimension of M and the Riemannian distance from x to the boundary of D.
Furthermore, Bismut type arguments have been used to derive results on short time asymptotics of the heat kernel by Malliavin and Stroock ([M-St]). Earlier Norris ([No]) studied stochastic formulae for heat semigroups on vector bundles over a compact Riemannian manifold for derivatives of arbitrary order and also gave applications on short time behaviour.

The scope of the present work is – in addition to provide a rather comprehensive description and discussion how to prove general gradient formulae – to transfer and extend the methods of Bismut, Elworthy/Li and Thalmaier to obtain stochastic representations of the Hessian

of harmonic functions and diffusion semigroups both in the non-intrinsic and the intrinsic case as well as to give some applications of fundamental type.
Already [E-L 1] contains a Hessian formula for compact manifolds (which follows as a special case from our results) but still non-intrinsic and with the disadvantages explained above.
But even if one generalizes Elworthy and Li's formula for the Hessian by introducing two finite energy processes (for more flexibility) analogously to the gradient case, it is not obvious how to pass over to the intrinsic context, because filtering out redundant noise – in mathematical terms taking conditional expectation with respect to the smaller filtration – is a linear operation, whereas the Hessian (and thus its stochastic representation) depends bilinearly on two entries that "contain noise". It turns out that one has to start out by an intrinsic martingale containing the gradient and carry out a covariant derivation that leads to a non-intrinsic second derivative formula of which the conditional expectation can be computed. The crucial point is to find an explicit expression for the noise-filtered version of the covariant derivative of the Dohrn-Guerra transport along the paths of the diffusion.
The resulting representation formula naturally includes derivatives of curvature terms since this transport itself was obtained by a pathwise equation based on the Ricci curvature on M. In fact, we could carry out these calculations by using a commutation formula of Arnaudon and Thalmaier, cf. [A-Th 4].
For some first corollaries, we show that the intrinsic Hessian formula substantially simplifies in the Ricci-parallel situation (i.e. $\nabla \operatorname{Ric} \equiv 0$). As well, there is a result for the Hessian of the logarithmic heat kernel $\operatorname{Hess}_x \log p(t, \cdot, y)$ involving expectations with respect to the Brownian bridge measure analogously to the earliest gradient formula of $\operatorname{grad}_x \log p(t, \cdot, y) = \frac{\operatorname{grad}_x p(t, \cdot, y)}{p(t,x,y)}$ by Bismut himself.

Following the ideas of Thalmaier and Wang, we then apply our formulae to prove Hessian estimates of the form

$$|\operatorname{Hess}_x u| \leq C\|u\|_D \tag{11}$$

for harmonic functions on a regular domain $D \subset M$ as well as similar results for the heat kernel or semigroup case. The main work from the Hessian representation formula to the final (deterministic) Hessian estimate (with explicit constants) consists of the construction of two finite energy processes on D varying on disjoint time intervals. The constants are then computed by comparing the Laplacian on D with that of a model manifold of suitable curvature.
Finally, we also give a proof of a pointwise gradient and Hessian estimate for positive harmonic functions in the case of a rotationally symmetric manifold.

To end this general description we should point out that our topic not only concerns questions and problems of differential geometry or stochastic analysis in the manifold-valued context, but also has direct applications in Euclidean matters. For instance, consider equation (1) on $\mathbb{R}^n$ with coefficient σ such that the associated second order differential operator (3) is (uniformly) elliptic. Then σ induces a Riemannian metric on $\mathbb{R}^n$ and together with the geometry coming from this metric we obtain important tools for the treatment of the partial differential operator. For example, questions of stability of expressions of the type $\mathbb{E}[f(X_t(x))]$ under perturbations of the initial value x, i.e. the question how such perturbations propagate in time, turn out to depend on the sign of the Ricci curvature corresponding to the induced metric.

The work is organized as follows.
Chapter 1 provides some basic notions and definitions of the theory of diffusions and stochastic flows on a finite-dimensional manifold. This part is kept very concise since nowadays there are some excellent and thorough introductions available on textbook level, cf. [St 2], [Hs 2], [H-Th] (in German) and for the theory of stochastic flows, although mainly treating the Euclidean case, the monograph by Kunita ([Ku]).
The second half of Chapter 1 gives typical representatives of space-time-harmonic functions. The main example of course is the *minimal heat semigroup* on M applied to a bounded measurable function, or, equivalently, the smooth *heat kernel* of this semigroup. We also present the martingale method that our differentiation of heat semigroups and similar objects is based on. The main argument makes use of the fact that if a family of real local martingales $M_s(\varepsilon)$ depends C^1 on a real parameter ε at 0 in the sense that $\frac{1}{\varepsilon}(M_s(\varepsilon) - M_s(0))$ converges uniformly on compact time intervals in probability, then the limit $\frac{\partial}{\partial\varepsilon}\big|_{\varepsilon=0} M_s(\varepsilon)$ is again a local martingale. All details concerning the involved *topology of semimartingales* on M (including non-trivial lifetime) can be found in [A-Th 1].

Chapter 2 first collects basic results about the derivative process associated to the diffusion which naturally appears when differentiating the martingales of Chapter 1. We study the defining stochastic differential equations for the derivative process in Stratonovich as well as in Itô form, since the latter one explicitly shows the relation of the formulae to curvature terms.
Whereas the derivative process mainly occurs in non-intrinsic representation formulae since it is in general not adapted to the filtration generated by $X_\bullet(x)$ (its differential at x depends on the behaviour of $X_\bullet(\,\cdot\,)$ in a neighbourhood of x), we can obtain intrinsic versions in terms of the Le Jan-Watanabe connection on M by filtering out redundant noise. Besides the precise meaning of these notions we also give an explicit argument, based on the Weitzenböck decomposition of the Laplacian on 1-forms, for the fact that the differentiated heat semigroup martingale preserves the (local) martingale property when the derivative process is replaced by the (noise-filtered) damped transport.
After these more advanced results, we return to basic methods and compute a first order martingale and the resulting stochastic representation formulae for the gradient of diffusion semigroups. We present a very short and immediate integration by parts proof as well as the Girsanov perturbation argument due to Thalmaier already mentioned above. Our most general Hessian martingale is obtained by this (second) method afterwards.
Moreover, we briefly outline the construction of the involved finite energy process, which is given in detail in the applications of Chapter 4 below.
The chapter finishes with intrinsic results derived by taking conditional expectation with respect to the smaller filtration.

Chapter 3 consists of new results for second order derivatives. We start out by the perturbation argument for the diffusion disturbed by two (small) real parameters and use Girsanov compensation to obtain a general, though non-intrinsic, Hessian (local) martingale depending on two finite energy processes. It turns out that if these processes are assumed to decay to zero on disjoint time intervals then after taking expectations many

terms of the quite complex structure cancel out and one gets a result comparable to that of Elworthy and Li. Analogously to the gradient case, Theorem 3.4 gives the precise assumptions under which our representation formula for the Hessian holds.
Since the general Hessian martingale, however, is not suitable for filtering out redundant noise, we revisit the proof in a special case and add an integration by parts procedure which admits to pass over to the intrinsic case.
As explained above, the most important step to achieve an intrinsic formula is to apply some commutation formula from [A-Th 4]. The final Hessian representation formula appears in Theorem 3.12. For technical reasons, these results are only proven for $L = \frac{1}{2}\Delta$ (without additional vector field), which refers to $X_{\bullet}(x)$ being a Brownian motion. Corollaries for the Hessian of the logarithmic heat kernel as well as for the Ricci parallel case follow more or less immediately.

The final Chapter 4 treats – as an important class of applications – gradient and Hessian estimates of (L-)harmonic functions and diffusion semigroups. The basic results are of Harnack type, but we give several modifications as well. In the positive harmonic case for instance, the estimates in some way can be compared with analytic results due to Cheng and Yau (cf. [Ch-Y] and [Sch]).
The first section herein presents the results for the gradient when using bounds of the space-time-harmonic functions and estimating the L^2-norm of the stochastic integral in the representation formula by an explicit construction of the finite energy process involving a quite subtle time change and careful Gronwall arguments.
We then carry over these ideas to our Hessian formula (again for $L = \frac{1}{2}\Delta$), which requires to construct two finite energy processes, being spatially separated with respect to different domains in addition to varying at disjoint times. Iteration of Gronwall, Burkholder-Gundy and Cauchy-Schwartz estimates then yield a comparable result for the Hessian, which has slightly different formulations for the harmonic function and the heat semigroup case.
As a concluding example for a more special situation, where the local estimates can be replaced by pointwise ones, we consider a rotationally symmetric manifold, the centre of symmetry being given by the initial value x of the diffusion.

To finish this introduction, I would like to express my gratitude to several people.
First of all, I should mention Prof. Dr. Anton Thalmaier, who turned my interest towards his work on Bismut formulae in the gradient case, encouraged me to try on the second derivative case and supported the whole work until its accomplishment.
Secondly, I thank Prof. Dr. Wolfgang Hackenbroch for his constant advice and interest in the subject as well as for the opportunity to present my progress in his seminar.
Another important person who supported me by answering many questions via e-mail, provided some important ideas and proofs (which are explicitly marked in the text) and read my drafts on the Hessian estimate in Chapter 4 very carefully, is Prof. Marc Arnaudon from the University of Poitiers, France.
Moreover, I have to thank Dr. Robert Denk, Dr. Stefan Bechtluft-Sachs, Dr. Ulrich Riegel and Michaela Lautenschlager for their advice throughout several periods of the last more than three years.
Finally, I would like to point out my gratitude to my fiancée and colleague Dipl.-Math. Stefanie Ulsamer, who carefully read the manuscript, made me become aware of some facts on differential geometry and supported me in many other aspects.

Chapter 1

Stochastic access to harmonic functions and diffusion semigroups on manifolds

In this basic chapter we briefly describe the situation and definitions that we start from when using stochastic methods to derive formulae for derivatives of heat semigroups on Riemannian manifolds. Most parts of the work are based on and can be understood with the knowledge of the Euclidean framework of stochastic analysis in addition to the facts about stochastic processes on Riemannian manifolds given below.
However, we occasionally sketch arguments of more geometrical type to introduce some well-known results. With regard to this and for a complete and thorough introduction to the theory of stochastic processes and analysis on manifolds we refer to textbooks like [H-Th], [Em 1] and, more recently, [St 2] and [Hs 2].

1.1. Geometric prerequisites

Throughout the whole work let M be a smooth n-dimensional manifold. We require M to have a countable basis for the topology, be connected, and to have (unless otherwise stated) no boundary. As usual, we assume manifolds, bundles and related objects to be smooth in the sense of C^∞, although most statements would hold in the C^2- or C^3-case as well.
In general, M is not assumed to be compact. Let $TM \xrightarrow{\pi} M$ denote the tangent bundle over M.
We assume that M is equipped with a Riemannian metric g, the metric could also be induced by a Whitney embedding and the bundle homomorphism defined subsequently, see Remark 1.1 below.
For some integer m we consider a homomorphism of vector bundles over M

$$A : M \times \mathbb{R}^m \to TM,$$

i.e. for fixed $x \in M$ the mapping $A(x) \equiv A(x, \cdot\,) : \mathbb{R}^m \to T_xM$ is linear and on the other hand for fixed $e \in \mathbb{R}^m$ we have the vector field $A(\cdot)e \equiv A(\cdot, e)$. If $(e_i)_{i=1,\ldots,m}$ denotes the standard orthonormal basis in $\mathbb{R}^m$, we write $A_i := A(\cdot)e_i$ for brevity.

Throughout this work we consider second order differential operators on M that can be represented in Hörmander form as

$$L = A_0 + \frac{1}{2}\sum_{i=1}^{m} A_i^2, \tag{1.1}$$

where A_0 is some vector field. We exclusively treat the elliptic case, which is the most basic one, i.e.

$$L = \frac{1}{2}\Delta_M + V, \tag{1.2}$$

where Δ_M denotes the Laplace-Beltrami operator on M and V is a vector field depending on the coefficients A and A_0.
For A this means that the pointwise adjoint homomorphism

$$A(x)^* : T_xM \to \mathbb{R}^m$$

is an isometric inclusion for each $x \in M$. In particular, $m \geq n$.
The explicit relation between V and A, A_0 is given by

$$V(x) = \frac{1}{2}\operatorname{trace} \nabla A(A(x)\,\cdot\,)(\,\cdot\,) + A_0(x) \tag{1.3}$$

(cf. [H-Th], Bem. 7.112), where in this case ∇ denotes the connection on $\operatorname{Hom}(\mathbb{R}^m, TM)$ induced by the Levi-Civita connection on M and thus $\nabla A : TM \times \mathbb{R}^m \to TM$ is a bilinear morphism.

There are several points of view that we can start from to derive this situation, which we discuss now briefly.

Remark 1.1. i) If we are only given some finite dimensional manifold M, we choose an arbitrary embedding $\iota : M \to \mathbb{R}^\ell$ of M into Euclidean space. Then we define $A(x)$ as the orthogonal projection $\mathbb{R}^\ell \to T_xM$ such that $A(x)A(x)^* = \mathrm{id}_{T_xM}$. Let now g be the metric induced by the Euclidean one on $\mathbb{R}^\ell$, i.e. $g(u,v) := \langle A(x)^*u, A(x)^*v\rangle_{\mathbb{R}^\ell}$ for arbitrary $u, v \in T_xM$.
This definition of g also works if a bundle homomorphism A is already given such that $A(x) : \mathbb{R}^m \to T_xM$ is surjective for each $x \in M$, and only the metric has to be chosen.
ii) On the other hand, if M is already provided with some Riemannian metric g, according to Nash's theorem ([Na]) we find an isometric embedding ι of (M, g) into Euclidean space of sufficiently high dimension and by this isometry we have the canonical choice of $A(x)^* = (d\iota)_x$ which determines A.
iii) Both of the upper situations are special cases of "gradient Brownian systems with drift", given if $A(x)^*$ is an isometric immersion (cf. [E-LJ-L], Example 1B).

In addition to these conventions, we will later sometimes use the Le Jan-Watanabe connection (cf. [E-LJ-L]) instead of the Levi-Civita connection, which implies

$$\operatorname{trace} \nabla A \otimes A = 0.$$

We give an introduction to these notions in Section 2.1.2 below.

1.2. Stochastic prerequisites

Although we are interested in heat semigroups related to the elliptic generator L given by (1.2), we translate it to terms of A and A_0 from the preceding section in order to determine the associated diffusion process via the Stratonovich equation

$$\mathrm{d}X = A(X) * \mathrm{d}Z + A_0(X)\mathrm{d}t. \tag{1.4}$$

Herein Z denotes an $\mathbb{R}^m$-valued Brownian motion on a filtered probability space

$$(\Omega; \mathscr{F}; \mathbb{P}; (\mathscr{F}_t)_{t\in\mathbb{R}_+}),$$

for which we adopt the following conventions:
The filtration $(\mathscr{F}_t^Z)$ is assumed to be complete and right continuous (which are also called the usual conditions) except for those cases where Girsanov techniques are used.
Then the assumptions have to be relaxed to local completeness where besides right continuity we only demand that $\mathscr{F}_0^Z$ already contains the σ-ideal

$$\mathscr{N} := \left\{ N \subset \bigcup_1^\infty N_i : N_i \in \bigcup_{s\geq 0} \mathscr{F}_s, \mathbb{P}(N_i) = 0 \right\}$$

(cf. [H-Th], p.250).
From standard theory of stochastic differential equations on manifolds we know that there is a partial solution flow $(X_t(x), \zeta(x))_{x\in M}$ to (1.4) in the sense that for $x \in M$ fixed, $X_t(x)$ is the strong solution to the stochastic differential equation defined on the stochastic interval $[0, \zeta(x)[$ starting in $X_0(x) = x$ with lifetime $\zeta(x)$. In this context, $\zeta(x)$ is a predictable stopping time for which a.s. on $\{\zeta(x) < \infty\}$ one has $X_t(x) \to \infty$ with $t \nearrow \zeta(x)$ in the one-point-compactification $\widehat{M} := M \cup \{\infty\}$ of M.
If we have a look on the sets

$$M_t(\omega) := \{x \in M : t < \zeta(x, \omega)\}$$

of starting points x where the solution path $X_{\bullet}(x, \omega)$ is still alive at time t, the solution flow has the following properties (see [Th 1] for this formulation and [Ku] for proofs):

i) For each $t \geq 0$, $\zeta(\,\cdot\,, \omega)$ is lower semicontinuous on M and therefore $M_t(\omega)$ an open subset of M.

ii) $X_t(\,\cdot\,, \omega)$ is a diffeomorphism from $M_t(\omega)$ onto an open subset of M.

iii) For each $t \geq 0$ the map $s \mapsto X_s(\,\cdot\,, \omega)$ from $[0, t]$ into $C^\infty(M_t(\omega), M)$ (endowed with its C^∞-topology) is continuous.

1.3. Space-time-harmonic functions and heat semigroups

In spite of the fact that we sometimes emphasize the heat semigroup case, the methods we develop in the following are not restricted to the computation of derivatives of heat semigroups. In fact, they will work for objects of type

$$\mathbb{E}\left[F_\tau \circ X_\tau(x)\, 1_{\{\tau<\zeta(x)\}}\right], \quad [0, \tau] \subset I, \tag{1.5}$$

($I = [0,t]$ or $I = \mathbb{R}_+$, τ a predictable stopping time), such that

$$F : I \times M \to \mathbb{R}, \qquad F_s = F(s, \cdot\,)$$

has the following properties:

i) F is C^1 with respect to the first and C^2 with respect to the second argument.

ii) Both the spatial differential $(s,x) \mapsto (dF_s)_x$ and the Hessian $(s,x) \mapsto (\nabla dF_s)_x$ of F are jointly continuous on $I \times M$.

iii) $F_s \circ X_s(x)$ yields a real local martingale for all $x \in M$, $0 \le s < \tau$.

By introducing an additional stopping time, all assumptions could be restricted to hold on an open subset D of M.

Remark 1.2. Condition iii) holds true, if $t \in \overline{\mathbb{R}}_+$ such that $\tau < t$ and $F : [0,t[\times M \to \mathbb{R}$ is a space-time-harmonic function with respect to L, i.e. F satisfies

$$\left(\frac{\partial}{\partial s} + L\right) F = 0 \qquad \text{on } \;]0,t[\times M. \tag{1.6}$$

Proof. This is a standard application of the Itô formula: According to the time dependence of F_s, one has to turn over to the semimartingale $(s, X_s)_{s\in\mathbb{R}_+}$ on the product manifold $\mathbb{R} \times M$, which yields

$$\mathrm{d}(F_s \circ X_s) = (dF_s)_{X_s} A(X_s)\mathrm{d}Z_s + \frac{\partial}{\partial s} F_s(X_s)\mathrm{d}s + LF_s(X_s)\mathrm{d}s \overset{m}{=} 0$$

due to F being space-time-harmonic (we write $\overset{m}{=}$ if the terms only differ by the differential of some local martingale). Recall that for $f \in C^2(M)$ and X_s solution to (1.4) we have $\mathrm{d}(f \circ X_s) = (df)_{X_s} A(X_s)\mathrm{d}Z_s + Lf(X_s)\mathrm{d}s$, which characterizes X_s as a diffusion process with respect to the (elliptic) generator L. □

Most of the applications will be either one of the following two cases:

i) (Dirichlet problem, stochastic representation of L-harmonic functions).
Consider $D \subset M$ open and bounded, $\tau(x) \equiv \tau_D(x)$ the first exit time of $X_s(x)$ from (the interior of) D, and $F_s \equiv u$ (not depending on s), where u is an L-harmonic function, i.e. $Lu = 0$.

   As a special case we could consider the Poisson kernel $p(\,\cdot\,, y)$ on a regular domain $D \subset M$ (i.e. D is open, relatively compact with smooth boundary) with respect to a fixed $y \in \partial D$.

ii) (Stochastic representation of diffusion semigroups).
Here we take a finite time horizon $I = [0,t]$ and $F_s = P_{t-s}f$, where f is bounded and measurable and P_s denotes the minimal heat semigroup on bounded measurable functions on M associated to L. We will give an explicit definition of this below. The inversion in the time index of the semigroup has to be made to switch from the heat equation $\frac{\partial}{\partial s} - L = 0$ to the space-time-harmonic context $\frac{\partial}{\partial s} + L = 0$.

Occasionally, we also speak of diffusion semigroups to emphasize that the expression "heat semigroup" does not only concern the particular case of $L = \frac{1}{2}\Delta$.

Another important example is the smooth heat kernel $F_s := p(t-s, \cdot, y)$ itself (again $y \in M$ fixed), which appears in the definition of the minimal heat semigroup below as well.

Note that both for the semigroup and the kernel, assumption ii) above has to be weakened since F_t then equals $P_0 f = f$ or some Dirac distribution, which fails to be differentiable with respect to x. Hence ii) holds true on $[0, t[\times M$ only.

We now give a brief description of the analytic semigroup theory our methods are related to, mainly for the sake of completeness, because we will only deal with stochastic representations of semigroups afterwards. According to our assumptions, we constrain this to the case of elliptic generators of the form $L = \frac{1}{2}\Delta_M + V$.

Definition 1.3 (Minimal heat semigroup). We call a family of linear operators on $L^\infty(M)$ (with identification of functions which differ only on a set of volume measure 0)

$$P_t : L^\infty(M) \to L^\infty(M), \quad t \in \mathbb{R}_+$$

a semigroup with generator $L = \frac{1}{2}\Delta_M + V$, if:

i) $P_0 f = f$, $P_{s+t} f = P_s \circ P_t f$, $f \in L^\infty(M)$, $s, t \geq 0$ (semigroup property).

ii) P_t preserves positivity ($P_t f \geq 0$, if $f \geq 0$) and is contractive ($\|P_t f\|_\infty \leq \|f\|_\infty$ or, equivalently, $P_t 1 \leq 1$) for all $t \geq 0$.

iii) For all test functions $f \in C_c^\infty(M)$ we have $\lim_{t\to 0} \frac{P_t f(x) - f(x)}{t} = Lf(x)$.

We call $(P_t)_{t\geq 0}$ minimal, if:

iv) For all $(Q_t)_{t\in\mathbb{R}_+}$ with the upper properties, we have $P_t f \leq Q_t f$ a.e. for all $f \in L^\infty(M)$ with $f \geq 0$ and all $t \geq 0$.

Well known from the theory of parabolic partial differential equations such a semigroup $(P_t)_{t\in\mathbb{R}_+}$ owns a smooth heat kernel $p \in C^\infty(]0, \infty[\times M \times M)$ which provides

$$(P_t f)(x) = \int_M p(t, x, y) f(y) \operatorname{vol}(\mathrm{d}y), \quad f \in L^\infty(M),\ x \in M,\ t > 0,$$

and $u(t, x) := P_t f(x)$ solves the heat equation

$$\begin{cases} \left(\frac{\partial}{\partial t} - L\right) u(t, x) = 0, \\ u(0, \cdot\,) = f. \end{cases} \tag{1.7}$$

From this property one can immediately deduce what is usually called *elliptic regularity* or *smoothing property* of the semigroup: For arbitrary $t > 0$, $P_t f \in C^\infty(M)$ even if f is just bounded and measurable.
Standard elliptic theory guarantees the existence and uniqueness of a minimal heat semigroup. In analytic terms, for $f \in L^2(M)$ the semigroup $P_t f$ can be constructed via the spectral theorem as $e^{t\widetilde{L}} f$, where $\widetilde{L}$ denotes the Friedrichs extension of $L|_{C_c^\infty(M)}$, see [D-Th], Cor. B.5. We can deduce the following statement from this result as well:

Theorem 1.4. *Let (M, g) be a Riemannian manifold and $(P_t)_{t \in \mathbb{R}_+}$ the minimal semigroup on $L^\infty(M)$ generated by the elliptic operator L as above. X_t shall denote the diffusion process with lifetime ζ and generator $L = \frac{1}{2}\Delta + V$. Then we have*

$$(P_t f)(x) = \mathbb{E}\left[(f \circ X_t(x)) 1_{\{t < \zeta(x)\}}\right], \quad f \in L^\infty(M). \tag{1.8}$$

See also [H-Th], 7.252.

Convention 1.5. For the purposes of this work take equation (1.8) as the definition of $P_t f$. Occasionally, we write *stochastic heat semigroup* to take this fact into account.

1.4. Martingale methods to differentiate diffusion semigroups

In [E-L 1] the authors give various stochastic representation formulae for first and second order derivatives of the minimal heat semigroup on $L^\infty(M)$. Their general way to obtain these results is to differentiate the right hand side of (1.8) under the expectation. The first disadvantage of this method is that f therefore has to be bC^1 (bC^2 respectively), i.e. f bounded and continuously differentiable of first or second order with bounded derivatives. Note, however, that derivatives of $P_t f$, $t > 0$, can be taken for all bounded measurable f because of the smoothing property of the semigroup, which coincides with the fact that Elworthy and Li's final formulae do not contain any derivatives of f. Indeed, the methods presented in this work are completely independent of any regularity of f exceeding measurability and boundedness.
Moreover, formal differentiation under the integral sign in (1.8) requires $\zeta \equiv \infty$ a.s., which means the driving diffusion is non-explosive. In fact, taking derivatives under the expectation corresponds to the associated semigroup on 1-forms

$$P_t^{(1)}(\alpha) := \mathbb{E}[X_t^* \alpha], \qquad \alpha \in \Gamma(T^*M), \tag{1.9}$$

where $X_t^*\alpha$ denotes the pullback of α by $X_t : M \to M$. In our case we have $\alpha = df$ which for $v \in T_x M$ yields

$$P_t^{(1)}(df)_x v = \mathbb{E}[(df)_{X_t(x)}(T_x X_t)v].$$

Herein $T_x X_t : T_x M \to T_{X_t(x)} M$ is the differential of $X_t(\,\cdot\,)$ at x which is well defined for all ω with $x \in M_t(\omega)$. A more thorough introduction of this process will be given at the beginning of the following chapter.
If we generalize this equation for explosive systems by

$$P_t^{(1)}(df)_x v := \mathbb{E}[(df)_{X_t(x)}(T_x X_t)v\, 1_{\{t < \zeta(x)\}}],$$

then $d(P_t f) = P_t^{(1)}(df)$ will not hold in general. For example, if $f \equiv 1$ on M, we find $P_t^{(1)}(df) = 0$, but $P_t f(x) = \mathbb{P}\{\zeta(x) > t\}$ and this value will depend on x in the case of non-trivial lifetime.

A more promising strategy to differentiate heat semigroups is to make use of the local martingale property of $P_{t-s} f(X_s(x))$, $0 \le s \le t$. As we discuss subsequently, derivatives of C^1-families of local martingales are again local martingales. Hence in the next chapter we will vary x for example along a smooth curve on M, differentiate at x in direction

v, and find that the resulting process is a true martingale under certain boundedness assumptions. By comparing expectations at time 0 and t this provides an expression for $(dP_tf)_xv$.
The whole next chapter will deal with first order derivatives calculated by modifications of this procedure.

We conclude this chapter by giving basic results about the differentiation of local martingales. For this purpose we sketch two notions of topologies on spaces of continuous M-valued processes. We refer to [A-Th 1] for proofs and further details.
Let $\mathcal{J}$ denote the set of predictable stopping times on the given filtered probability space. For $\xi \in \mathcal{J}$ and $F \subset M$ some closed subset, we write $D_c(F;\xi)$ for the set of continuous adapted F-valued processes with lifetime ξ and $\mathcal{S}(F;\xi)$ for the set of continuous F-valued semimartingales. We define

$$\widehat{D}_c(F) := \bigcup_{\xi\in\mathcal{J}} D_c(F;\xi) \quad \text{and} \quad \widehat{\mathcal{S}}(F) := \bigcup_{\xi\in\mathcal{J}} \mathcal{S}(F;\xi).$$

In [A-Th 1] the authors give a base of neighbourhoods of $X \in \widehat{D}_c(F)$ ($\widehat{\mathcal{S}}(F)$ respectively), which defines a separated topology on this space, called the *topology of compact convergence in probability* (the *topology of semimartingales* respectively). In fact this is first done for $F = \mathbb{R}^n$ and then transferred to the general case by the use of a smooth proper embedding into $\mathbb{R}^n$; it turns out that the definition does not depend on the choice of the embedding.
We have $\widehat{\mathcal{S}}(F) \subset \widehat{D}_c(F)$ and convergence for the semimartingale topology implies compact convergence in probability. The opposite implication does not hold for an arbitrary sequence of semimartingales, however.
As was proven in [A-Th 1], Proposition 2.10, the topology of semimartingales coincides with the topology of compact convergence in probability, if we restrict ourselves to the closed subspace $\widehat{\mathcal{M}}_\nabla(M)$ of $\widehat{D}_c(M)$ which consists of all M-valued ∇-martingales (with lifetime ζ). Recall that a stochastic process $(X_t)_{t\in[0,\zeta[}$ taking values in M is a ∇-martingale if for any $f \in C^\infty(M)$ the real process $f\circ X - \int \nabla df(\mathrm{d}X,\mathrm{d}X)$ is a local martingale. Herein $\int \nabla df(\mathrm{d}X,\mathrm{d}X)$ denotes the so-called df-quadratic variation defined in [H-Th], 7.58.
Thus for sequences in $\widehat{\mathcal{M}}_\nabla(M)$ we obtain:

Theorem 1.6. *If a sequence in $\widehat{\mathcal{M}}_\nabla(M)$ converges uniformly on compact sets in probability, i.e. with respect to the topology on $\widehat{D}_c(M)$, its limit is again a ∇-martingale and convergence takes place in the sense of semimartingales.*

Proof. See [Em 2], Thm. 4.11. □

Since in the real case the notion of ∇-martingales corresponds to that of local martingales, we will make use of this result in the following version.

Corollary 1.7. *Let M be a smooth manifold. If $(m_s(x))_{s\in\mathbb{R}_+,\, x\in M}$ is a family of real local martingales that depends C^1 on $x \in M$ (with respect to compact convergence in probability) then the differential $(dm_s)_x$ yields again a local martingale, but now taking values in T_x^*M. (Equivalently: $(dm_s)_xv$ is a real local martingale for any $v \in T_xM$.)*

Regularity results on solutions of stochastic differential equations are discussed in [A-Th 1], part 3. Actually, our diffusions $(X_s(x))_{s\in\mathbb{R}_+}$ depend C^1 on the initial value x and the same holds for the so-called *derivative process* $(T_xX_s \equiv d(X_s)_x)_{s\in\mathbb{R}_+}$. We make use of the latter fact when computing second order derivatives.
Hence, in our situation the real local martingale $(F_s \circ X_s(x))_{s\in\mathbb{R}_+}$ is C^1 with respect to $x \in M$ and Corollary 1.7 can be applied.

Chapter 2

Differentiation of diffusion semigroups

The intention of this chapter is to present the differentiation method for diffusion semigroups and related objects which our approach is based on. We discuss two slightly different methods in the case of gradient formulae: on the one hand a variation of a one-parameter family of real martingales (derived from a perturbation argument including Girsanov compensation), which was introduced by [Th 1], and on the other hand a combination of a directional derivative of a real martingale smoothly depending on $x \in M$ and an integration by parts argument. In the case of first derivatives the latter one is the easier method and can be found e.g. in [A-Th 3].
The main theorem establishes a stochastic representation formula depending on a suitable finite energy process K_s taking values in the Euclidean space T_xM for fixed $x \in M$. We will give an explicit choice of such a K_s, which in the domain case is a non-trivial problem. Finally, we show how our results can be converted to an *intrinsic* representation theorem, i.e. a formula that only relies on the geometrical data of the manifold but is independent of the particular choice of the stochastic differential equation defining our diffusion X_s.
To this end, we need the notion of the derivative process of X_s, two related filtrations that the diffusion is adapted to, and a linear connection on TM that may differ from the Levi-Civita connection, but naturally appears in the intrinsic case.

Throughout this and the following chapters $F : I \times M \to \mathbb{R}$ will denote a function satisfying conditions i) - iii) stated right before Remark 1.2. For brevity, however slightly inaccurate, we refer to this by saying "F is space-time-harmonic".

2.1. Redundant noise: Source and filtering

2.1.1. The derivative process

As already mentioned above, our main tool for computations will be the (local) martingale property of first order derivatives of $F_s(X_s(\,\cdot\,))$. Simple differentiation at $x \in M$ in direction $v \in T_xM$ leads to the local martingale

$$(dF_s)_{X_s(x)} T_x X_s v, \tag{2.1}$$

where T_xX_s denotes the differential of $X_s(\,\cdot\,)$ at x, which exists for all $\omega \in \Omega$ with $x \in M_s(\omega)$ and solves the formally differentiated version of the stochastic differential equation (1.4)

$$DV_s = (\nabla A)V_s * \mathrm{d}Z_s + (\nabla A_0)V_s \mathrm{d}s, \quad V_0 = v$$

(for an introduction to these facts see e.g. [El 4], §8).
The symbol ∇ in this context denotes the Levi-Civita connection on M and the induced connection on vector bundles generated by TM (such as tensor bundles, etc.), respectively. Note the difference in the notation of the process V_s on TM in contrast to the vector field V of (1.3) that appears in the generator L of X_s (and hence in the Itô equation for $DV_s = DTX_s$ below).
DV_s stands for the covariant Itô differential V_s with values in $\mathrm{Hom}(TM, X^*TM)$, explained below (in fact, this is only true if $X_{\bullet}(x)$ has lifetime $\zeta(x) = \infty$ a.s., otherwise, according to Section 1.2, one has to write $V_s : TM_s \to X^*TM$). Recall that $V_s(x) : T_xM \to T_{X_s(x)}M$ is a linear map for each $x \in M_s$.
We write $//_{0,s}$ for the stochastic parallel translation $T_xM \to T_{X_s(x)}M$ along the paths of $X_s(x)$, see Definition 2.9 below.
In terms of this transport the covariant Itô differential $DV_s(x)$ is defined as

$$DV_s(x) := //_{0,s}\, \mathrm{d}\left(//_{0,s}^{-1}\, V_s(x)\right)$$

where the differential on the right hand side is the usual Itô differential for a process taking values in the fixed Euclidean space T_xM.

As a solution to a stochastic differential equation with continuous driving process (and time independent coefficients) T_xX_s possesses a time continuous version, and, moreover, because of the diffeomorphic property of $X_s(\,\cdot\,)$ on M_s every T_xX_s is bijective and its inverse process $(T_xX_s)^{-1} \in \mathrm{Hom}_x(X^*TM, TM)$ has continuous paths (a.s.) as well. One also refers to TX as the *derivative process* (or, as some authors emphasize the flow property of solutions, the *derivative flow*) associated to X.

Remark 2.1. We rewrite the defining stochastic differential equation for TX_s in terms of Itô differentials for a better motivation of the following arguments. The Stratonovich equation reads

$$DTX_s = \nabla_{TX_s} A * \mathrm{d}Z_s + \nabla_{TX_s} A_0 \mathrm{d}s \tag{2.2}$$

whereas the Itô version is given by

$$\begin{aligned} DTX_s &= \nabla_{TX_s} A\, \mathrm{d}Z_s - \frac{1}{2} R(TX_s, \mathrm{d}X_s)\, \mathrm{d}X_s + \nabla_{TX_s} V\, \mathrm{d}s \\ &= \nabla_{TX_s} A\, \mathrm{d}Z_s - \frac{1}{2} \mathrm{Ric}^\sharp(TX_s)\, \mathrm{d}s + \nabla_{TX_s} V\, \mathrm{d}s. \end{aligned} \tag{2.3}$$

Herein $V = \frac{1}{2} \operatorname{trace} \nabla_A A + A_0 = \frac{1}{2} \sum_{i=1}^m \nabla_{A_i} A_i + A_0$, cf. (1.3). The right hand side involves the Riemannian curvature tensor $R \in \Gamma(T^*M \otimes T^*M \otimes T^*M \otimes TM)$, see for instance [Pe], 2.2.1, and $R(X,Y)Z \equiv R(X,Y,Z)$ by convention for vector fields X, Y, Z. The Ricci curvature $\mathrm{Ric} \in \Gamma(T^*M \otimes T^*M)$ is defined by $\mathrm{Ric}(X,Y) := \operatorname{trace}\langle R(X,\,\cdot\,)\,\cdot\,, Y\rangle$. Thus reading the 1-form $\mathrm{Ric}(X,\,\cdot\,)$ as a vector field $\mathrm{Ric}^\sharp(X)$ according to $\langle \mathrm{Ric}^\sharp(X), Y\rangle = \mathrm{Ric}(X,Y)$, we are given $\mathrm{Ric}^\sharp(X) = \operatorname{trace} R(X,\,\cdot\,)(\,\cdot\,)$.

Equation (2.3) is found e.g. in [A-Th 2], Example 4.10. In that paper the authors make use of the theory of complete lifts to carry out the formal differentiation of the basic Itô-stochastic differential equation.
A direct proof of (2.3) from (2.2) requires to compute

$$DTX_s = \nabla_{TX_s} A \, \mathrm{d}Z_s + \frac{1}{2}\sum_{i=1}^m \nabla_{A_i}\nabla_{TX_s} A_i \, \mathrm{d}s - \frac{1}{2}\sum_{i=1}^m \nabla_{TX_s}\nabla_{A_i} A_i \, \mathrm{d}s + \nabla_{TX_s} V \mathrm{d}s$$

by verifying that $\sum_{i=1}^m \nabla_{A_i}\nabla_{TX_s} A_i \, \mathrm{d}s$ is the covariation of $\nabla_{TX_s} A$ and Z_s. It remains to insert $R(X,Y,Z) = \nabla_X\nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z$ and to observe that the term depending on the Lie bracket $[TX_s, A_i]$ equals zero (which relies on the fact that TX_s is the spatial derivative of X_s and A the leading coefficient in the time development of X_s).
One should notice that (2.3) holds ([A-Th 2], 4.10) for any torsion-free connection ∇ on M. This plays an important role if we want to use the Le Jan-Watanabe connection in some cases on which we focus next.

2.1.2. The Le Jan-Watanabe connection

A brief outline of the following can be found in [A-Th 4] subsequently to Definition 7.5 (note that by assumption $\mathrm{im}(A) = TM$). For a thorough treatment we refer to [E-LJ-L].

Definition 2.2 (The Le Jan-Watanabe connection).
The Le Jan-Watanabe covariant derivative ∇^{LJW} on TM ($= \mathrm{im}(A)$) is defined by the equation

$$\nabla_v^{\mathrm{LJW}} Z = A(x)\, d\big(A^*(\,\cdot\,)Z(\,\cdot\,)\big)_x(v), \tag{2.4}$$

where $x \in M$, $v \in T_xM$ and $Z \in \Gamma(TM)$ ($= \Gamma(\mathrm{im}(A))$).

Since $A^*Z : M \to \mathbb{R}^m$ is a smooth map, the composition $A\, d(A^*Z)$ yields a smooth section in $\mathrm{Hom}(TM, TM)$. The further properties of a linear connection on TM are verified easily.

A trivial, but important consequence of the definition is that $\nabla = \nabla^{\mathrm{LJW}}$ satisfies the following property:

Remark 2.3. Whenever $x \in M$, $e \in (\ker A(x))^\perp$, $v \in T_xM$, we have $\nabla_v A(\,\cdot\,)e = 0$.

This follows by applying the Leibniz rule for ∇^{LJW} to the identity $A(A^*A) = (AA^*)A = A$, which yields

$$\begin{aligned}\nabla_v^{\mathrm{LJW}} A(\,\cdot\,)e &= (\nabla_v^{\mathrm{LJW}} A)(A^*(x)A(x))e + A(x)d(A^*(\,\cdot\,)A(\,\cdot\,)e)_x v \\ &= \nabla_v^{\mathrm{LJW}} A(\,\cdot\,)e + \nabla_v^{\mathrm{LJW}} A(\,\cdot\,)e,\end{aligned}$$

as $A^*(x)A(x)$ is the orthogonal projection from $\mathbb{R}^m$ to $(\ker A(x))^\perp$.

This last property of ∇^{LJW} can be reformulated as follows: There always is a decomposition $K(\,\cdot\,) \oplus K(\,\cdot\,)^\perp$ of the trivial bundle $M \times \mathbb{R}^m$ over M such that $A(\,\cdot\,)k = 0 \in TM$ for $k \in K(\,\cdot\,)$ and $\nabla A(\,\cdot\,)\ell = 0 \in \mathrm{Hom}(TM, TM)$ for $\ell \in K(\,\cdot\,)^\perp$. In particular, we have

$$\mathrm{trace}(\nabla A \otimes A) = 0 \qquad (\in \Gamma(\mathrm{Hom}(TM, TM) \otimes TM))\,,$$

where $\nabla A \otimes A \in \Gamma(\mathrm{Bil}(\mathbb{R}^m, \mathbb{R}^m; \mathrm{Hom}(TM, TM) \otimes TM))$.

We emphasize the content of Example 1B in [E-LJ-L]:

Proposition 2.4. *If M is isometrically immersed in $\mathbb{R}^m$ and $A(x)$ the orthogonal projection on T_xM (i.e. we have a gradient Brownian system) as in particular in the first two cases of Remark 1.1, then the Le Jan-Watanabe and the Levi-Civita connection coincide.*

Convention 2.5. Unless otherwise stated, we refer to the Levi-Civita connection and simply write ∇.
However, in some cases we also make use of Remark 2.3 and may call it *Le Jan-Watanabe property* (of the connection). Mainly, this will occur in the following chapters which treat *intrinsic* formulae. In that cases the proofs hold if we either assume that we are given a gradient Brownian system (where both connections coincide), or, in the more general situation, we consider ∇^{LJW} instead of ∇ without carrying the index LJW along the computations. Then, all appearing differential operators, particularly all Laplacians, have to be taken with respect to ∇^{LJW}.
We point out that the gradient system case is adequate to derive all our applications.

2.1.3. Comparison of two filtrations

Now let $(\mathscr{F}_s^Z)_{s\in\mathbb{R}_+}$ be the complete and right continuous filtration generated by the driving Brownian motion Z, in contrast to $\big(\mathscr{F}_s^{X(x)}\big)_{s\in\mathbb{R}_+}$ corresponding to our diffusion process X starting from some given $x \in M$. In general, $\mathscr{F}^{X(x)}$ will be smaller than $\mathscr{F}^Z$. In fact the difference "increases" with that of the dimensions m of Z and n of our manifold M (for example, let $M = \mathbb{R}^n$ and A be an orthogonal projection $\mathbb{R}^m \to \mathbb{R}^n$, $A_0 = -\frac{1}{2}\operatorname{trace}\nabla_A A$ such that $V = 0$, then $X_{\boldsymbol{\cdot}}(x)$ is a n-dimensional Brownian motion and $\mathscr{F}^Z$ is generated by X and an independent $(m-n)$-dimensional Brownian motion $\widetilde{X}$).
Note that when dealing with the two real martingales $F_s \circ X_s(x)$ and $(dF_s)_{X_s(x)}T_xX_sv$, the first one is adapted to $\big(\mathscr{F}_s^{X(x)}\big)$ whereas the differentiated process is only $\mathscr{F}^Z$-adapted because in general T_xX_s is not measurable with respect to $\mathscr{F}_s^{X(x)}$ since the derivative depends on the (stochastic) germ of $X_s(\,\cdot\,)$ at x and not only on $X_s(x)$ itself. Hence the terms containing TX carry "redundant noise" which is not intrinsic in geometric terms. The appropriate solution to this problem is known as the method of "filtering out redundant noise". We will give the main result and refer to [E-Y] for proofs. For this purpose it is necessary to introduce the notions of the orthonormal frame bundle and horizontal lifts.

Definition 2.6. With respect to our Riemannian manifold (M,g) we have the *orthonormal frame bundle* $O(M) \xrightarrow{\pi} M$ given by

$$O(M) := \dot{\bigcup_{x\in M}} P_x, \qquad P_x := \pi^{-1}x := \{u : \mathbb{R}^n \to T_xM \mid u \text{ an isometry}\}.$$

Remark 2.7. i) We identify $u \in P_x$ with $(u_1,\dots,u_n) := (ue_1,\dots,ue_n)$ which is an orthonormal basis of T_xM. Herein $(e_1,\dots,e_n)$ denotes the standard basis of $\mathbb{R}^n$.

ii) $O(M) \xrightarrow{\pi} M$ owns the structure of a *principal bundle* with *structure group* $O(n)$, the group of orthogonal transformations of $\mathbb{R}^n$; cf. [H-Th], Def. 7.121. In particular, $O(M)$ is again a smooth manifold.

Now the Levi-Civita connection ∇ on TM induces pointwise – i.e. for each $u \in O(M)$ – a decomposition

$$T_u(O(M)) = V_u \oplus H_u$$

consisting of the *vertical part*

$$V_u := \{v \in T_uO(M) : (d\pi)v = 0\} \equiv \ker(d\pi)$$

(which is canonical and does not depend on the choice of ∇) and a *horizontal part* H_u (invariant under right action of the group $O(n)$) constructed the following way.
For $u \in O(M)$ with $\pi(u) = x$ let

$$\begin{aligned} H_u := \{(d\hat{Y})_x v : v \in T_xM, \hat{Y} \in \Gamma(O(M)/U) \text{ with } \hat{Y}(x) = u \\ \text{and } \nabla_v \hat{Y} := (\nabla_v(\hat{Y}e_1), \ldots, \nabla_v(\hat{Y}e_n)) = 0\}, \end{aligned}$$

where $U \subset M$ ist an open set containing x.
The mapping $h : \pi^*TM \cong H \hookrightarrow T(O(M))$ with $h_u : T_{\pi(u)}M \xrightarrow{\cong} H_u$ is then called the *horizontal lift* (of the $O(n)$-connection induced by ∇) and provides the *standard-horizontal vector fields* $L_1, \ldots, L_n \in \Gamma(T(O(M)))$ given by $L_i(u) := h_u(ue_i)$.
As a consequence, we have the *horizontal Laplacian* $\Delta^{\text{hor}} := \sum_{i=1}^n L_i^2$ on $O(M)$. For details of all these facts and definitions, see [H-Th], p.415 ff.
Occasionally, it is useful to work rather with stochastic processes taking values in $O(M)$ than directly with those on M. For this reason we need an adequate way to lift a M-valued semimartingale up to $O(M)$.

Definition 2.8. Let X denote a semimartingale taking values in M. An $O(M)$-valued semimartingale U is called a *horizontal lift* of X, if

i) $\pi \circ U = X$ a.s.

ii) $\int_U \omega = 0$, i.e. U is *horizontal*.

Herein $\omega \in \Gamma(T^*O(M) \otimes T_{\text{id}}O(n))$ is the so-called *connection form* (cf. [H-Th], 7.127) and $\int_U \omega$ the *Stratonovich-integral of* ω *along* U ([H-Th], 7.63).

According to [H-Th], Satz 7.141, for a given random $u_0 \in O_{X_0}(M)$ there exists a unique horizontal lift U of X onto $O(M)$ with $U_0 = u_0$ a.s.
As an immediate consequence, we now can give the explicit construction of the *stochastic parallel transport* along X:

Definition 2.9. Let U denote a horizontal lift of X. Then the family of isometries

$$//_{0,s} := U_s \circ U_0^{-1} \in \Gamma(T^*M \otimes X_s^*TM), \quad s < \zeta,$$

is called *stochastic parallel transport* along (the paths of) X. This definition does not depend on the choice of the horizontal lift.

Now we return to our initial problem of the two different filtrations.

Let $(U_s(x))_{s\in\mathbb{R}_+}$ denote a horizontal lift of $(X_s(x))_{s\in\mathbb{R}_+}$ to the orthonormal frame bundle $\pi : O(M) \to M$ starting at some $U_0 \in \pi^{-1}(X_0(x))$. We can "reduce" our driving Brownian motion on $\mathbb{R}^m$ to one on $\mathbb{R}^n$ (but depending on a chosen initial value $x \in M$ of the diffusion) by the stochastic integral

$$\widetilde{B}_s := \int_0^s U_r^{-1} A(X_r(x)) \mathrm{d}Z_r \tag{2.5}$$

and another one on T_xM of the same dimension (but independent of the choice of the horizontal lift) by the isometric transformation

$$B_s := U_0 \widetilde{B}_s = \int_0^s U_0 U_r^{-1} A(X_r(x)) \mathrm{d}Z_r = \int_0^s //_{0,r}^{-1} A(X_r(x)) \mathrm{d}Z_r, \tag{2.6}$$

which turns out to be the martingale part of the stochastic anti-development of $X_{\boldsymbol{\cdot}}(x)$ given by the stochastic integral $\int_0^{\boldsymbol{\cdot}} //_{0,r}^{-1} * \mathrm{d}X_r$ (cf. [A-Th 3], formula (3.3) and [H-Th] for the general theory of anti-development).
Then the following equalities hold

$$\mathscr{F}^B = \mathscr{F}^{\widetilde{B}} = \mathscr{F}^{X(x)} = \mathscr{F}^U \tag{2.7}$$

(cf. [E-Y], p.161).

Now *filtering out redundant noise* from the derivative flow can be interpreted as taking conditional expectation of TX_s with respect to $\mathscr{F}_s^{X(x)}$ by the following result.

Proposition 2.10. *We consider the case of a gradient Brownian system with drift V (or more generally, $\nabla = \nabla^{\mathrm{LJW}}$ and consequently* $\operatorname{trace} \nabla A \otimes A = 0$*, cf. Convention 2.5).*
For each $x \in M$ we have a linear transport $W_s : T_xM \to T_{X_s(x)}M$ defined by the (pathwise) covariant linear equation along $X_{\boldsymbol{\cdot}}(x)$

$$\frac{\nabla}{\partial s} W_s = -\frac{1}{2} \operatorname{Ric}(W_s, \cdot\,)^\sharp + \nabla V(W_s), \quad W_0 = \mathrm{id}_{T_xM}\,. \tag{2.8}$$

Then W_s a.s. equals the conditional expectation with respect to $\big(\mathscr{F}_s^{X(x)}\big)$ of the derivative process along the paths of $X_s(x)$, i.e.

$$\mathbb{E}[T_xX_s | \mathscr{F}_s^{X(x)}] = W_s, \tag{2.9}$$

where the left hand side by definition means

$$\mathbb{E}[T_xX_s | \mathscr{F}_s^{X(x)}] := //_{0,s} \mathbb{E}[//_{0,s}^{-1} T_xX_s | \mathscr{F}_s^{X(x)}]$$

(and herein the conditional expectation is taken on the fixed Euclidean space T_xM).

Proof. This can be found in [E-Y], 3 A Thm. A (the compactness assumption on M is for simplicity only and can be omitted). □

W_s is often referred to as the *damped* or *Dohrn-Guerra* transport along the paths of $X_s(x)$ (the latter name is according to both authors work about geodesic deviation in stochastic mechanics).
We give a brief and heuristic argument how (2.8) can be derived from (2.3) and the property of Remark 2.3 according to the Le Jan-Watanabe connection.

Consider the orthogonal decomposition of the trivial bundle over M given by $M \times \mathbb{R}^m = M \times \big(\ker A(\,\cdot\,) \oplus (\ker A(\,\cdot\,))^\perp\big)$. Consequently, along the paths of $X_s(x)$ we can write $Z_s = Z_s^0 + Z_s^\perp$, where $Z_s^0 \in \ker A(X_s(x))$ and $Z_s^\perp \in (\ker A(X_s(x)))^\perp$. Then the defining equation for $X_s(x)$ in in the Itô sense reads

$$\mathrm{d}X_s(x) = A(X_s(x))\mathrm{d}Z_s^0 + A(X_s(x))\mathrm{d}Z_s^\perp + \text{ drift}$$

and (2.3)

$$\begin{aligned} DTX_s(x) =& \nabla_{TX_s(x)}A(X_s(x))\mathrm{d}Z_s^0 + \nabla_{TX_s(x)}A(X_s(x))\mathrm{d}Z_s^\perp \\ & - \frac{1}{2}\operatorname{Ric}^\sharp(TX_s(x))\mathrm{d}s + \nabla_{TX_s(x)}V(X_s(x))\mathrm{d}s. \end{aligned}$$

In the first equation the term $A(X_s(x))\mathrm{d}Z_s^0$ equals 0 due to the definition of Z_s^0. Thus, taking conditional expectation with respect to $\mathscr{F}_s^{X(x)}$ means to drop the terms driven by Z_s^0 which cancels the first part on the right hand side of the second equation when we pass over from TX_s to W_s. But according to the property of Remark 2.3 $\nabla A(X_s(x)) = 0$ on $(\ker A(X_s(x)))^\perp$, so the other term depending on ∇A vanishes as well and we end up with (2.8).
The rigorous deduction of these facts is done via horizontal lifts of the processes up to the orthonormal frame bundle and can be found in [E-Y].

Equation (2.8) defining W_s can also be rewritten as one in T_xM by using stochastic parallel translation $//_{0,s}$ along the paths of $X_s(x)$ and replacing W_s by $Q_s := //_{0,s}^{-1} W_s$. If we define $\operatorname{Ric}_{//_{0,s}} := //_{0,s}^{-1}\operatorname{Ric}^\sharp_{X_s(x)}//_{0,s} : T_xM \to T_xM$ as well as $(\nabla V)_{//_{0,s}} := //_{0,s}^{-1}(\nabla V)_{X_s(x)}//_{0,s} : T_xM \to T_xM$, we get from (2.8)

$$\frac{\partial}{\partial s}Q_s = -\frac{1}{2}\operatorname{Ric}_{//_{0,s}} Q_s + (\nabla V)_{//_{0,s}}Q_s, \quad Q_0 = \mathrm{id}_{T_xM}\,. \tag{2.10}$$

Remark 2.11. According to this for $0 \neq v \in T_xM$ the paths of Q_sv solve a linear differential equation on T_xM with respect to time, starting in $v \neq 0$. Hence the solution to this linear equation has no zeroes, and consequently Q_s (a.s.) remains an isomorphism on T_xM for all $0 \leq s < \zeta(x)$. As W_s is derived from Q_s by an isometric transformation the same statement holds for W_s (except for its range being $T_{X_s(x)}M$).
So throughout the following chapters on Hessian formulae we will often use the *inverse damped transport* $W_s^{-1} \in \Gamma(\mathrm{Hom}(X_s^*TM, TM))$.

We conclude this discussion by a direct proof via the *Weitzenböck decomposition theorem* for the Laplacian on 1-forms that $(dF_s)_{X_s(x)}W_sv \equiv (dF_s)_{X_s(x)}//_{0,s}Q_sv$ is a local martingale for every $v \in T_xM$ (cf. [Th 2], Lemma 2.1). Obviously, it is adapted with respect to $(\mathscr{F}^{X(x)})$ because the only stochastic part in (2.10) ((2.8) respectively) is given by stochastic parallel translation $//_{0,s}$ along $X_s(x)$.

Theorem 2.12. *If $(Q_s)_{s\in\mathbb{R}_+}$ is a solution to (2.10) and $v \in T_xM$, then $(dF_s)_{X_s(x)}//_{0,s}Q_sv$ is a local martingale.*

Proof. For the proof we have to lift things up to the orthonormal frame bundle $O(M)$. For $u \in O(M)$ let $h_u : T_{\pi(u)}M \to T_uO(M)$ denote the horizontal lift (of the $O(n)$-connection

induced by ∇) as above. Besides the horizontal vector fields $L_1, \ldots, L_n \in \Gamma(TO(M))$ we will also use the *horizontal lift* $\overline{V} \in \Gamma(TO(M))$ of V determined by $\overline{V}_u = h_u(V_{\pi(u)})$.
Let U be a an $O(M)$-valued diffusion starting in $U_0 = u_0 \in O(M)_x$ with $\pi \circ U = X$ and generator $\frac{1}{2}\Delta^{\mathrm{hor}} + \overline{V} = \frac{1}{2}\sum_{i=1}^n L_i^2 + \overline{V}$. As one easily verifies, the horizontal Laplacian satisfies $\Delta^{\mathrm{hor}}(f \circ \pi) = (\Delta f) \circ \pi$ for every smooth function f on M.
Moreover, we identify the set of 1-forms $\Gamma(T^*M)$ with the set of *equivariant functions* $\{f : O(M) \to \mathbb{R}^n : f(ug) = g^{-1}f(u), u \in O(M), g \in O(n)\}$. The identification is done by $\alpha \mapsto f_\alpha$, where $f_\alpha : O(M) \to \mathbb{R}^n$, $f_\alpha^i(u) := \alpha_{\pi(u)}(ue_i)$, $1 \le i \le n$.
Recall that the Laplacian $\Delta^{(1)}$ on 1-forms is well defined by

$$\Delta^{(1)} df = d\Delta f, \quad f \in C^\infty(M).$$

Then, the *Weitzenböck decomposition theorem* (cf. [Jo], Thm. 3.3.3) tells us that for all $\alpha \in \Gamma(T^*M)$ we have

$$\Delta^{(1)}\alpha = \Box\alpha - \mathrm{Ric}(\alpha^\sharp, \cdot\,), \tag{2.11}$$

where $\Box$ denotes the so called *rough Laplacian* on 1-forms given by $\Box\alpha = \operatorname{trace}\nabla^2\alpha$ which in detail means

$$\Box : \Gamma(T^*M) \xrightarrow{\nabla^{T^*M}} \Gamma(T^*M \otimes T^*M) \xrightarrow{\nabla^{T^*M\otimes T^*M}} \Gamma(T^*M \otimes T^*M \otimes T^*M) \xrightarrow{\text{trace}} \Gamma(T^*M)$$

(last trace taken with respect to the first two entries of the triple tensor product).
As a consequence, $f_{\Box\alpha} = \Delta^{\mathrm{hor}} f_\alpha$.

Now we can write $n_s := (dF_s)_{X_s(x)}//_{0,s}Q_sv = \langle f_{dF_s}(U_s), U_0^{-1}Q_sv\rangle_{\mathbb{R}^n}$. As $U_0^{-1}Q_sv$ has paths with bounded variation because of (2.10), Itô's product formula yields

$$\begin{aligned} \mathrm{d}n_s &= \langle \mathrm{d}(f_{dF_s}(U_s)), U_0^{-1}Q_sv\rangle + \langle f_{dF_s}(U_s), \mathrm{d}(U_0^{-1}Q_sv)\rangle \\ &= \langle \mathrm{d}(f_{dF_s}(U_s)), U_0^{-1}Q_sv\rangle + \langle f_{dF_s}(U_s), U_0^{-1}(-\tfrac{1}{2}\mathrm{Ric}_{//_{0,s}} Q_s + (\nabla V)_{//_{0,s}} Q_s)v\,\mathrm{d}s\rangle. \end{aligned}$$

On the other hand, by the diffusion property of U we have (modulo differentials of local martingales)

$$\begin{aligned} \mathrm{d}(f_{dF_s}(U_s)) &\overset{m}{=} \frac{\partial}{\partial s} f_{dF_s}(U_s)\mathrm{d}s + \tfrac{1}{2}\Delta^{\mathrm{hor}} f_{dF_s}(U_s)\mathrm{d}s + \overline{V} f_{dF_s}(U_s)\mathrm{d}s \\ &= \left[-\tfrac{1}{2} f_{\Delta^{(1)}dF_s}(U_s) - f_{d(dF_sV)}(U_s) + \tfrac{1}{2} f_{\Box dF_s}(U_s) + \overline{V} f_{dF_s}(U_s)\right]\mathrm{d}s \end{aligned}$$

where we have used $\frac{\partial}{\partial s} f_{dF_s} = f_{d\frac{\partial}{\partial s}F_s}$ and, because F_s is space-time- harmonic, $d\frac{\partial}{\partial s}F_s = -d(\frac{1}{2}\Delta F_s + VF_s) = -\frac{1}{2}\Delta^{(1)}dF_s - d(dF_sV)$. $\overline{V} \in \Gamma(TO(M))$ herein needs to be read as a derivation on $C^\infty(O(M))$ ($C^\infty(O(M), \mathbb{R}^n)$ respectively).
Now putting all this together, we get

$$\begin{aligned} \mathrm{d}n_s &\overset{m}{=} -\tfrac{1}{2}(\Delta^{(1)}dF_s)_{X_s(x)}//_{0,s}Q_sv\,\mathrm{d}s + \tfrac{1}{2}(\Box dF_s)_{X_s(x)}//_{0,s}Q_sv\,\mathrm{d}s \\ &\quad - d(dF_sV)//_{0,s}Q_sv\,\mathrm{d}s + \langle \overline{V} f_{dF_s}(U_s), U_0^{-1}Q_sv\rangle \mathrm{d}s \\ &\quad - \tfrac{1}{2}(dF_s)_{X_s(x)}//_{0,s}\mathrm{Ric}_{//_{0,s}} Q_sv\,\mathrm{d}s + (dF_s)_{X_s(x)}//_{0,s}(\nabla V)_{//_{0,s}} Q_sv\,\mathrm{d}s \\ &= 0. \end{aligned}$$

Herein we have used

$$(dF_s)_{X_s(x)}//_{0,s}\,\mathrm{Ric}_{//_{0,s}}\,Q_s v = (dF_s)_{X_s(x)}\,\mathrm{Ric}^\sharp_{X_s(x)}\,//_{0,s}Q_s v = \mathrm{Ric}_{X_s(x)}((dF_s)^\sharp, //_{0,s}Q_s v)$$

to apply the Weitzenböck formula as well as (cf. [H-Th], Satz 7.133)

$$\overline{V}_{U_s}(f_{dF_s})^i = (\nabla_V dF_s)_{X_s(x)}(U_s e_i) \equiv (\nabla dF_s)(V(X_s(x)), U_s e_i)$$

and the chain (or product) rule

$$d(dF_s V)_{X_s(x)} = (\nabla dF_s)(V(X_s(x)), \,\cdot\,) + (dF_s)_{X_s(x)}\nabla V.$$

□

2.2. General methods and non-intrinsic first order derivatives

Remark 2.13. In the situation that we call *non-intrinsic*, all statements or assumptions concerning measurability or adaptedness are meant with respect to the (larger) filtration $(\mathscr{F}^Z_s)$.

As we described before, our first order derivative formulae make use of the fact that the local martingale of the following theorem, which depends on the choice of some T_xM-valued adapted process K with absolutely continuous paths, is a true martingale if K is a *finite energy process*:

For any interval I of the type $[0,t]$ or $[0,t[$ and a finite dimensional vector space E we denote by

$$\mathbb{H}(I,E) := \{\gamma : I \to E \text{ absolutely continuous}, \|\dot\gamma\| \in L^2(I, \mathrm{d}s)\}$$

the *Cameron-Martin space* of curves taking values in E. We write $\mathbb{H}_0(I,E)$ if we only want to consider curves with $\gamma(0) = 0$.

Bounded adapted processes with sample paths in these spaces are called *finite energy processes* (or, to be more exactly, L^p-finite energy processes), if they fulfill $\left(\int_0^t \|\dot K_s\|^2\,\mathrm{d}s\right)^{1/2} \in L^p(\mathbb{P})$, $p > 0$. We will always deal with cases where $p = 1+\alpha$ for some $\alpha > 0$.

Moreover, we will now write X_{s*} instead of TX_s which emphasizes the role of the differential as a push forward and lets the formulae become a bit more readable.

Theorem 2.14. *Let $(K_s)_{s\in[0,t]}$ be an adapted process with sample paths in $\mathbb{H}([0,t], T_xM)$ and $K_0 = v \in T_xM$. Then $(n_s)_{0\le s\le \sigma\wedge t}$ given by*

$$n_s := (dF_s)_{X_s(x)} X_{s*} K_s - F_s(X_s(x)) \int_0^s \langle X_{r*}\dot K_r, A(X_r(x))\mathrm{d}Z_r\rangle \tag{2.12}$$

is a (real) local martingale, where $\sigma < \zeta(x)$ is a stopping time and hence $F_s \circ X_s(x)$ is a well-defined local martingale for $0 \le s \le \sigma$.

There are several possibilities to prove this theorem which we will now study and compare in detail because afterwards these methods shall be modified to compute second order derivatives. At the end of this paragraph we will exploit the result of the theorem for a heat semigroup derivative formula. But first, we give the following little application that will be useful afterwards.

Theorem 2.15 (Integration by parts formula).
Let $t > 0$, F be bounded on $[0,t] \times M$ and X non-explosive. Then the equation

$$\mathbb{E}[(dF_t)_{X_t(x)} X_{t*} K_t] = \mathbb{E}\left[F_t \circ X_t(x) \int_0^t \langle X_{r*}\dot{K}_r, A(X_r(x))\mathrm{d}Z_r\rangle\right] \tag{2.13}$$

holds for any bounded adapted K with sample paths in $\mathbb{H}_0([0,t], T_xM)$, if additionally $\mathbb{E}[\sup_{0\le s\le t} |(dF_s)_{X_s(x)} X_{s*} K_s|] < \infty$ and $\int_0^s \langle X_{r*}\dot{K}_r, A(X_r(x))\mathrm{d}Z_r\rangle$ is a martingale on $[0,t]$.

Proof. From Theorem 2.14 we know that

$$n_s := (dF_s)_{X_s(x)} X_{s*} K_s - F_s(X_s(x)) \int_0^s \langle X_{r*}\dot{K}_r, A(X_r(x))\mathrm{d}Z_r\rangle$$

is a local martingale on $[0,t]$ starting in 0 because $K_0 = 0$. So the assertion follows from $\mathbb{E}n_t = \mathbb{E}n_0 = 0$ if n_s is a true martingale. This is the case if $\{n_\sigma : 0 \le \sigma \le t \text{ stopping time}\}$ is uniformly integrable. So we have to check this for the two terms on the right hand side of the upper equation separately. The second one can be estimated by an upper bound for $|F|$ times the integral which is a martingale and therefore has the uniform integrability property we need. For the first term it is already given by the assumed boundedness of the expectation of the maximum process. □

In the heat semigroup case the result for differentiable initial function $P_0 f = f$ is the following.

Corollary 2.16. *Consider $f \in bC^1(M)$ (i.e. a bounded C^1-function with bounded derivative), $t > 0$, X non-explosive and K as in the theorem. Then we have*

$$\mathbb{E}[(df)_{X_t(x)} X_{t*} K_t] = \mathbb{E}\left[f \circ X_t(x) \int_0^t \langle X_{r*}\dot{K}_r, A(X_r(x))\mathrm{d}Z_r\rangle\right], \tag{2.14}$$

under the assumption that $\mathbb{E}[\sup_{0\le s\le t} |d(P_{t-s}f)_{X_s(x)} X_{s*} K_s|] < \infty$ and the stochastic integral $\int_0^s \langle X_{r*}\dot{K}_r, A(X_r(x))\mathrm{d}Z_r\rangle$ is a martingale.

2.2.1. Integration by parts

Proof of Theorem 2.14 by an integration by parts argument. According to Section 1.4 we know that $m_s := (dF_s)_{X_s(x)} X_{s*}$ is a local martingale with values in $T_x^*M \equiv \mathrm{Lin}(T_xM, \mathbb{R})$. Integration by parts yields (recall that (K_s) has absolutely continuous paths which therefore are of bounded variation on $[0,t]$)

$$\mathrm{d}(m_s K_s) = (\mathrm{d}m_s)K_s + m_s \mathrm{d}K_s \overset{m}{=} m_s \mathrm{d}K_s = m_s \dot{K}_s \mathrm{d}s.$$

On the other hand, the geometrical version of Itô's formula tells us by the space-time-harmonicity of F (cf. Section 1.3) that $\mathrm{d}(F_s \circ X_s(x)) = (dF_s)_{X_s(x)} A(X_s(x))\mathrm{d}Z_s$ and so again by the product formula we compute

$$\mathrm{d}\left(F_s(X_s(x)) \int_0^s \langle X_{r*}\dot{K}_r, A(X_r(x))\mathrm{d}Z_r\rangle\right) \overset{m}{=} (dF_s)_{X_s(x)} X_{s*}\dot{K}_s \mathrm{d}s = m_s \dot{K}_s \mathrm{d}s,$$

because the (differential of the) covariation process of the two real local martingales appearing on the left hand side is the product of the Itô differentials $(dF_s)_{X_s(x)} A(X_s(x))\mathrm{d}Z_s$ and $\langle A(X_s(x))^* X_{s*}\dot{K}_s, \mathrm{d}Z_s\rangle$ and we can use that $AA^* = \mathrm{id}_{TM}$ and $\mathrm{d}Z_s^i \mathrm{d}Z_s^j = \delta_{ij}\mathrm{d}s$. As the right hand sides of the last two equations are equal, the theorem is proven. □

2.2.2. Perturbation with Girsanov compensation

A second possibility to prove Theorem 2.14 according to A. Thalmaier ([Th 1]) is the following:

Proof of Theorem 2.14 by perturbation with Girsanov compensation. We replace our initial stochastic differential equation (1.4) by one with a considerably small additional parameter ε such that

$$\mathrm{d}X_s^\varepsilon = A(X_s^\varepsilon) * \mathrm{d}Z_s^\varepsilon + A_0(X_s^\varepsilon)\mathrm{d}s \tag{2.15}$$

where the driving process $\mathrm{d}Z_s^\varepsilon = \mathrm{d}Z_s + \varepsilon k_s \mathrm{d}s$ is the original Brownian motion plus a linear perturbation by an adapted process k_s with values in $\mathbb{R}^m$ such that $\int_0^t |k_s|^2 \mathrm{d}s < \infty$ a.s. Without loss of generality, we assume that $k_s \in (\ker A(X_s(x))^\perp$; otherwise we replace k_s by $A(X_s(x))^* A(X_s(x))k_s$ because $A(X_s(x))^* A(X_s(x))$ is the orthogonal projection onto the complement of the kernel.
To derive this situation in a mathematically rigorous way let us define the process $H_s^\varepsilon(x)$, $x \in M$, by the following pathwise equation (we will choose a proper initial value H_0^ε afterwards):

$$\begin{aligned} \mathrm{d}H_s^\varepsilon &= X_{s*}^{-1} A(X_s \circ H_s^\varepsilon)\varepsilon k_s \mathrm{d}s, \\ H_s^0 &= \mathrm{id}_M \,. \end{aligned} \tag{2.16}$$

Then $H_s^\varepsilon(\,\cdot\,,\omega)$ will be a smooth mapping from an open subset of M to M. We now define our perturbed solution $X_s^\varepsilon(x)$ by

$$X_s^\varepsilon(x) := X_s \circ H_s^\varepsilon(x), \qquad x \in M. \tag{2.17}$$

Applying Itô's generalized formula, which tells us how to compute the Itô differential of a composition of a M-valued process with one depending smoothly on a parameter in M (cf. [Ku], Ch. 3.3, for the Euclidean case which is easily translated to the geometrical situation), we derive (2.15) by

$$\begin{aligned} \mathrm{d}X_s^\varepsilon &= A(X_s^\varepsilon) * \mathrm{d}Z_s + A_0(X_s^\varepsilon)\mathrm{d}s + X_{s*} * \mathrm{d}H_s^\varepsilon \\ &= A(X_s^\varepsilon) * (\mathrm{d}Z_s + \varepsilon k_s \mathrm{d}s) + A_0(X_s^\varepsilon)\mathrm{d}s. \end{aligned} \tag{2.18}$$

Since we have changed the driving process of the stochastic differential equation into Z_s^ε which is no longer a (local) martingale, $F_s \circ X_s^\varepsilon(x)$ will not have the local martingale property either. But this can be compensated by a proper change of the probability measure according to Girsanov's theorem.
If we define G_s^ε as the stochastic exponential

$$G_s^\varepsilon = \mathscr{E}\left(-\int_0^s \varepsilon\langle k_r, \mathrm{d}Z_r\rangle\right) = \exp\left(-\int_0^s \varepsilon\langle k_r, \mathrm{d}Z_r\rangle - \frac{1}{2}\int_0^s \varepsilon^2 \|k_r\|^2\,\mathrm{d}r\right), \tag{2.19}$$

we have a probability measure $\mathbb{Q}$ (depending on ε) which is locally equivalent to $\mathbb{P}$ given by the relation $\mathbb{Q}|\mathscr{F}_s^Z = G_{s\wedge\rho}^\varepsilon \mathbb{P}|\mathscr{F}_s^Z$ for any stopping time $0 \le \rho \le s \wedge t$ such that $G_{\cdot\wedge\rho}^\varepsilon$ is a true martingale. Z_s^ε now is a local $\mathbb{Q}$-martingale and therefore – according to Lévy's characterization theorem – a Brownian motion with respect to $\mathbb{Q}$ on $[0,\rho]$. This is immediately checked because the Girsanov theorem says that if $\mathbb{Q}$ has density $\mathscr{E}(O)$ with

respect to $\mathbb{P}$ for a real local martingale O and $Z = (Z^i)$, $1 \le i \le m$, is a $\mathbb{R}^m$-valued local martingale with respect to $\mathbb{P}$, then $Z^i - [Z^i, O]$ has the same property with respect to $\mathbb{Q}$. Now, by pathwise uniqueness of solutions to our initial stochastic differential equation (1.4), we conclude that as $F_s \circ X_s(x)$ is a $\mathbb{P}$-local martingale, $(F_s \circ X_s^\varepsilon(x))$ is a $\mathbb{Q}$-local martingale and such is $(F_s \circ X_s^\varepsilon(x))G_s^\varepsilon$ with respect to $\mathbb{P}$ (at first for times $s \in [0, \rho]$, by choosing a suitable increasing sequence of stopping times the statement holds on $[0, \sigma \wedge t]$). Hence the latter one provides a C^1-family of local martingales parametrized by ε.
Differentiating at 0 then leads to the real local martingale

$$N_s := \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} (F_s \circ X_s^\varepsilon(x))G_s^\varepsilon = (dF_s)_{X_s(x)} \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} X_s^\varepsilon(x) + (F_s \circ X_s) \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} G_s^\varepsilon,$$

where the notation $\frac{\partial}{\partial\varepsilon}\big|_{\varepsilon=0}\gamma(\varepsilon) \in T_xM$ for a smooth curve $\gamma :]-\delta, \delta[\to M$, $\gamma(0) = x$, refers to the geometrical definition of the tangent space.
As G_s^ε is written as the exponential function of a linear plus a quadratic term of ε, we have

$$\frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} G_s^\varepsilon = -\int_0^s \langle k_r, \mathrm{d}Z_r\rangle,$$

so differentiation of (2.17) leads to

$$N_s = (dF_s)_{X_s(x)} X_{s*} \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} H_s^\varepsilon(x) - F_s \circ X_s(x) \int_0^s \langle k_r, \mathrm{d}Z_r\rangle. \tag{2.20}$$

We formally differentiate equation (2.16) for $\mathrm{d}H_s^\varepsilon$ (which contains a linear factor ε, so we can apply the product rule) with respect to ε and find after integration over $[0, s]$ that

$$\frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} H_s^\varepsilon(x) = \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} H_0^\varepsilon(x) + \int_0^s X_{r*}^{-1} A(X_r(x)) k_s \mathrm{d}s. \tag{2.21}$$

The mathematical justification of this procedure is given in a brief remark below. We now choose the smooth curve $\varepsilon \mapsto H_0^\varepsilon(x)$ such that $\frac{\partial}{\partial\varepsilon}\big|_{\varepsilon=0} H_0^\varepsilon(x) = v$ for a given vector $v \in T_xM$. For the sake of notational convenience we introduce

$$K_s := v + \int_0^s X_{r*}^{-1} A(X_r(x)) k_s \mathrm{d}s \tag{2.22}$$

which is an adapted process with paths in $\mathbb{H}([0, t], T_xM)$.
According to $k_s \in (\ker A(X_s(x)))^\perp$ and $A(X_s(x))^* A(X_s(x))$ being the orthogonal projection onto $(\ker A(X_s(x)))^\perp$ this immediately implies

$$k_s = A(X_s(x))^* X_{s*} \dot{K}_s, \tag{2.23}$$

so that if we start out by any given finite energy process K_s we can raise the last equation to the definition of k_s at the beginning of the proof.
Putting all this together we end up with the local martingale $N_s = (dF_s)_{X_s(x)} X_{s*} K_s - F_s \circ X_s(x) \int_0^s \langle X_{r*} \dot{K}_r, A(X_r(x)) \mathrm{d}Z_r\rangle$ and the theorem is proven. □

Remark 2.17. An in this case rather trivial remark which gets some significance in the second derivative case below is that we may have chosen $H_0^\varepsilon = \mathrm{id}_M$ and consequently $\frac{\partial}{\partial\varepsilon}\big|_{\varepsilon=0} H_0^\varepsilon(x) = 0$ to derive the theorem for $\widetilde{K}_s := K_s - v$ and add the local martingale (cf. the integration by parts proof) $(dF_s)_{X_s(x)} X_{s*} v$ afterwards.

Remark 2.18 *(On the formal differentiation of the defining equation for H_s^ε).*
By definition, (2.16) means that for any $f \in C_c^\infty(M)$ we have

$$f \circ H_s^\varepsilon(x) = f \circ H_0^\varepsilon(x) + \int_0^s (df)_{H_r^\varepsilon(x)} X_{r*}^{-1} A(X_r \circ H_r^\varepsilon(x)) \varepsilon k_r \mathrm{d}r.$$

Differentiating with respect to ε yields

$$\begin{aligned}(df)_x \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} & H_s^\varepsilon(x) \\ &= (df)_x \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} H_0^\varepsilon(x) + \int_0^s \frac{\nabla}{\partial\varepsilon}\Big|_{\varepsilon=0} \left[(df)_{H_r^\varepsilon(x)} X_{r*}^{-1} A(X_r \circ H_r^\varepsilon(x)) \varepsilon k_r\right] \mathrm{d}r\end{aligned}$$

(this is done path by path and one easily verifies that the derivative can be taken inside the integral sign). Herein the covariant differential $\frac{\nabla}{\partial\varepsilon}\big|_{\varepsilon=0} Y(\gamma(\varepsilon)) \in T_{\gamma(0)}M$ for a vector field $Y \in \Gamma(TM)$ and a smooth curve γ in M can be defined by $\frac{\partial}{\partial\varepsilon}\big|_{\varepsilon=0} L_{\gamma|[0,\varepsilon]}^{-1} Y(\gamma(\varepsilon))$ where L denotes the parallel transport induced by our connection ∇, and this equals $\nabla_{\dot\gamma(\varepsilon)} Y$.
As the composition of $(df)_{H_r^\varepsilon(x)} \circ X_{r*}^{-1} A(X_r \circ H_r^\varepsilon(x)) \varepsilon k_r$ can be read as a bilinear map $\mathrm{Hom}((H_r^\varepsilon)^* TM, \mathbb{R}) \times (H_r^\varepsilon)^* TM \to \mathbb{R}$ one uses the Leibniz rule to obtain

$$\begin{aligned}\frac{\nabla}{\partial\varepsilon}\Big|_{\varepsilon=0} & \left[(df)_{H_s^\varepsilon(x)} X_{s*}^{-1} A(X_s \circ H_s^\varepsilon(x)) \varepsilon k_s\right] \\ &= \nabla df \left(\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} H_s^\varepsilon(x), X_{s*}^{-1} A(X_s(x)) 0 k_s \right) + (df)_x \frac{\nabla}{\partial\varepsilon}\Big|_{\varepsilon=0} \left[X_{s*}^{-1} A(X_s \circ H_s^\varepsilon(x)) \varepsilon k_s\right]\end{aligned}$$

where the first term on the right hand side vanishes. Now using suitable functions f (e.g. with $(df)_x$ the coordinate projections of a fixed basis of T_xM) we derive

$$\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} H_s^\varepsilon(x) = \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} H_0^\varepsilon(x) + \int_0^s \frac{\nabla}{\partial\varepsilon}\Big|_{\varepsilon=0} \left[\varepsilon (X_{r*}^{-1} A(X_r \circ H_r^\varepsilon(x)) k_r)\right] \mathrm{d}r. \qquad (2.24)$$

Once more differentiating by the product rule proves (2.21) above.

2.2.3. Main result for the gradient

Prepared with the preceding computations, we are now able to state the application for first order derivatives of space-time-harmonic functions and a modification for heat semigroups due to the strong Markov property of $X_{\bullet}(x)$.
We treat the general case where X_t might be explosive.

Theorem 2.19 (Non-intrinsic representation formula for the gradient).
Let $\zeta(\,\cdot\,)$ be the lifetime of $X_t(\,\cdot\,)$. Fix $x \in M$ and let τ denote the first exit time of $X_{\bullet}(x)$ from a relatively compact neighbourhood D of x. Moreover, let $f : M \to \mathbb{R}$ be bounded measurable, $t > 0$, $v \in T_xM$ and $\tau_0 := \tau \wedge t$.
Then for every bounded adapted process $(K_s)_{s\in[0,t]}$ with sample paths in $\mathbb{H}([0,t], T_xM)$, which additionally fulfills $K_0 = v$, $K_s|_{s\in[\tau_0,t]} = 0$ and $(\int_0^t \|\dot K_s\|^2 \mathrm{d}s)^{1/2} \in L^{1+\alpha}(\mathbb{P})$ for some $\alpha > 0$, we have the following stochastic representation of $(dF_0)_x$:

$$\langle \mathrm{grad}_x F_0, v\rangle \equiv (dF_0)_x v = -\mathbb{E}\left[(F_{\tau_0} \circ X_{\tau_0}(x)) \int_0^{\tau_0} \langle X_{s*} \dot K_s, A(X_s(x)) \mathrm{d}Z_s\rangle\right]. \qquad (2.25)$$

In the heat semigroup case $F_s = P_{t-s}f$, $f \in L^\infty(M)$, we have the gradient representation formula

$$\langle \operatorname{grad}_x P_t f, v\rangle \equiv d(P_t f)_x v = -\mathbb{E}\left[(f\circ X_t(x))1_{\{t<\zeta(x)\}}\int_0^t \langle X_{s*}\dot K_s, A(X_s(x))\mathrm{d}Z_s\rangle\right] \quad (2.26)$$

with the convention that the integrand equals 0 *on paths where $t \geq \zeta(x)$.*

Proof. Obviously, $\tau_0 < \zeta(x)$ because $\tau < \zeta(x)$. Let n^{τ_0} denote the local martingale of Theorem 2.14 on $[0,t]$ stopped at τ_0. Our assumptions above assure that it is a true martingale:
To prove this we have to show that the family $\{n_\sigma^{\tau_0} : 0 \leq \sigma \leq t \text{ stopping time}\}$ of real random variables is uniformly integrable. Since F is continuous on the compact set $[0,t]\times \bar D$, choose $C_0 := \sup_{(s,y)\in[0,t]\times\bar D} |F_s(y)| < \infty$. Moreover, in Section 1.3 we have assumed that $(s,y)\mapsto (dF_s)_y$ is jointly continuous on $[0,t]\times M$, so we find $C_1 := \sup_{(s,y)\in[0,t]\times\bar D} \|d(F_s)_y\| < \infty$.
If C_2 is an upper bound for $\|K_s\|$, $0\leq s\leq t$, then obviously

$$|n_\sigma^{\tau_0}| \leq C_1 \sup_{[0,t]} \|T_x X_s\| C_2 + C_0 \left|\int_0^\sigma \langle A(X_s(x))^* X_{s*}\dot K_s, \mathrm{d}Z_s\rangle\right|. \quad (2.27)$$

As a standard result about the derivative process on a compact manifold one knows that $\sup_{[0,t]}\|T_x X_s\| \in L^p(\mathbb{P})$ for arbitrary $1\leq p<\infty$, cf. [Th 1] and [Li]. We need this sup only taken on $0\leq s\leq \tau_0$ (and modify M outside D if M is not compact such that T_xX_s for $s\leq\tau_0$ is not affected by this). Application of the Burkholder-Davis-Gundy and Hölder inequalities yields the existence of a $C<\infty$ such that for any $\sigma\leq\tau_0$ (recall that $A(X_s(x))^* : T_xX_s \to \mathbb{R}^m$ is an isometry)

$$\begin{aligned}\mathbb{E}\left|\int_0^\sigma \langle A(X_s(x))^* X_{s*}\dot K_s, \mathrm{d}Z_s\rangle\right| &\leq c\mathbb{E}\left(\int_0^{\tau_0} \|X_{s*}\dot K_s\|^2 \mathrm{d}s\right)^{\frac12}\\ &\leq c\left[\mathbb{E}\left(\sup_{0\leq s\leq\tau_0}\|X_{s*}\|\right)^{\frac{1+\alpha}{\alpha}}\right]^{\frac{\alpha}{1+\alpha}}\left[\mathbb{E}\left(\int_0^{\tau_0}\|\dot K_s\|^2\mathrm{d}s\right)^{\frac{1+\alpha}{2}}\right]^{\frac{1}{1+\alpha}} \leq C\end{aligned}$$

which proves the uniform integrability of $\{n_\sigma^{\tau_0} : 0\leq\sigma\leq t\}$.
Now comparison of the expectations of $n_0^{\tau_0}$ and $n_{\tau_0}^{\tau_0}$ yields (2.25) because of $X_{0*} = \mathrm{id}_{T_xM}$, $K_0 = v$ and $K_{\tau_0} = 0$.

In the heat semigroup case $F_s = P_{t-s}f$ we first assume that $K_s \equiv 0$ on $[t-\varepsilon, t]$ for some $\varepsilon > 0$. As above, we have $C_0 := \sup_{(s,y)\in[0,t]\times\bar D} |P_{t-s}f(y)| \leq \|f\|_\infty < \infty$, but for the differential only $C_1 := \sup_{(s,y)\in[0,t-\varepsilon]\times\bar D}\|d(P_{t-s}f)_y\| < \infty$ (due to elliptic regularity).
Hence reproducing (2.27), we verify that $n_s^{\tau\wedge(t-\varepsilon)}$ is a true martingale.
Thus from $\mathbb{E}n_0 = \mathbb{E}n_{\tau\wedge(t-\varepsilon)}$ we find that

$$d(P_tf)_xv = -\mathbb{E}\left[(P_{t-\tau\wedge(t-\varepsilon)}f\circ X_{\tau\wedge(t-\varepsilon)}(x))\int_0^{\tau\wedge(t-\varepsilon)}\langle X_{s*}\dot K_s, A(X_s(x))\mathrm{d}Z_s\rangle\right].$$

As $X_t(x)$ is the solution to (1.4), we can make use of its strong Markov property for the bounded functional $f\circ \mathrm{pr}_\sigma$ to obtain in the nonexplosive case

$$\mathbb{E}[(f\circ\mathrm{pr}_\sigma)\circ X_{(\tau\wedge(t-\varepsilon))+\boldsymbol{\cdot}}(x)|\mathscr{F}^Z_{(\tau\wedge(t-\varepsilon))}] = \mathbb{E}[(f\circ\mathrm{pr}_\sigma)\circ X_{\boldsymbol{\cdot}}(\,\cdot\,)]\circ X_{(\tau\wedge(t-\varepsilon))}(x)$$

which yields – taken at time $\sigma = t - (\tau \wedge (t-\varepsilon))$ – that $\left(P_{t-(\tau\wedge(t-\varepsilon))}f\right)\left(X_{(\tau\wedge(t-\varepsilon))}(x)\right) = \mathbb{E}[f \circ X_t(x)|\mathscr{F}^Z_{(\tau\wedge(t-\varepsilon))}]$. Hence the result for $d(P_tf)_xv$ above turns over to equality (2.26). In the explosive case, according to Theorem 1.4 we have

$$(P_tf)(x) = \mathbb{E}\left[(f \circ X_t(x))1_{\{t<\zeta(x)\}}\right].$$

So the strong Markov property now reads

$$\left(P_{t-(\tau\wedge(t-\varepsilon))}f\right)\left(X_{t-(\tau\wedge(t-\varepsilon))}(x)\right) = \mathbb{E}[f \circ X_t(x)1_{\{t<\zeta(x)\}}|\mathscr{F}^Z_{\tau_0}]$$

and besides this the proof works the same way.

In the case that we are given a general finite energy process K, we may approximate it by suitable K^ε with $K|_{[0,t-\varepsilon]} = K^\varepsilon|_{[0,t-\varepsilon]}$, but $K^\varepsilon|_{[t-\frac{\varepsilon}{2},t]} \equiv 0$ such that K^ε still matches the assumptions of the theorem. It remains to let $\varepsilon \to 0$ by dominated convergence using the finite energy condition for K_s. □

Remark 2.20. It is possible to replace the $L^{1+\alpha}$-assumption on the time integral over $\dot{K}$ by L^1 if one requires D being small enough such that some geometrical data about normed tangent vectors on D is bounded, cf. [Th 1], section 8. We will not make use of this result here.

From Theorem 2.19 we immediately derive the original Bismut formula (cf. [Bi]) for the gradient of the logarithm of the heat kernel associated to our driving diffusion X_s. This formula has been a great source of motivation for the development of the whole theory presented in this work.

Corollary 2.21. *Let* $p(\,\cdot\,,\,\cdot\,,\,\cdot\,) :]0,\infty[\times M \times M \to \mathbb{R}_+$ *denote the smooth (minimal) heat kernel associated to L, such that for any* $f \in L^\infty(M)$ *we have*

$$P_tf(x) = \mathbb{E}\left[(f \circ X_t(x)1_{\{t<\zeta(x)\}}\right] = \int_M p(t,x,y)f(y)\,\mathrm{vol}(\mathrm{d}y). \tag{2.28}$$

Then we can write the derivative of $\log p(t,\,\cdot\,,y)$ *in direction* $v \in T_xM$ *as a conditional expectation the following way:*

$$d(\log p(t,\,\cdot\,,y))_xv = \frac{d(p(t,\,\cdot\,,y))_xv}{p(t,x,y)} = \mathbb{E}\left[I_{\tau\wedge t}|X_t(x)=y\right], \tag{2.29}$$

where the martingale I_r, $0 \le r \le t$, *is given by*

$$I_r := -\int_0^r \langle X_{s*}\dot{K}_s, A(X_s(x))\mathrm{d}Z_s\rangle, \tag{2.30}$$

τ *and* K_s *fulfilling exactly the same conditions as in Theorem 2.19.*

Proof. This is shown by comparing the two representations

$$d(P_tf)_xv = \int_M d(p(t,\,\cdot\,,y))_xvf(y)\,\mathrm{vol}(\mathrm{d}y), \tag{2.31}$$

(which simply is (2.28) differentiated under the volume integral) and

$$d(P_tf)_xv = \int_M \mathbb{E}\left[I_{\tau\wedge t}|X_t(x)=y\right]p(t,x,y)f(y)\,\mathrm{vol}(\mathrm{d}y) \tag{2.32}$$

for test functions $f \in C_c(M)$.
The latter formula is obtained from (2.25) by the defining relation of the bridge measure for $A \in \mathscr{F}_s^{X(x)} \cap \{s < \zeta(x)\}$, $s < t$, as follows. We have

$$\mathbb{P}^t_{x,y}(A) = \int_A z_s \mathrm{d}\mathbb{P}, \quad z_s := \frac{p(t-s, X_s(x), y)}{p(t,x,y)}. \tag{2.33}$$

Due to the space-time-harmonicity of $(s,x) \mapsto p(t-s,x,y)$, the process z_s is a (strictly positive) martingale on $[0,t[$ (of L^1-norm 1), hence (2.33) is consistent with the inclusion $\mathscr{F}_r^{X(x)} \subset \mathscr{F}_s^{X(x)}$, $r < s$.
Note that $\mathbb{P}^t_{x,y}$ is concentrated on $\{t < \zeta(x)\}$: For $B \in \mathscr{B}(M)$ we have $X_s(x)^{-1}(B) \subset \mathscr{F}_s^{X(x)} \cap \{s < \zeta(x)\}$ and hence

$$\mathbb{P}|_{\mathscr{F}_s^{X(x)} \cap \{s<\zeta(x)\}} \circ X_s(x)^{-1}(B) = \mathbb{E}[1_B(X_s(x))1_{\{s<\zeta(x)\}}] = P_s 1_B(x) = \int_B p(s,x,y) \operatorname{vol}(\mathrm{d}y).$$

Consequently, for any $s < t$

$$\begin{aligned}\mathbb{P}^t_{x,y}\{s < \zeta(x)\} &= \mathbb{P}^t_{x,y}\{X_s(x) \in M, s < \zeta(x)\} = \mathbb{E}\left[1_M(X_s(x))1_{\{s<\zeta(x)\}} \frac{p(t-s, X_s(x), y)}{p(t,x,y)}\right] \\ &= \int_M \frac{p(s,x,\eta)p(t-s,\eta,y)}{p(t,x,y)} \operatorname{vol}(\mathrm{d}\eta) = 1,\end{aligned}$$

and passing over to the limit $s \nearrow t$ yields that $\mathbb{P}^t_{x,y}$ lives on the paths with no explosion up to time t.
We want to show

$$\mathbb{P}(A \cap \{s < \zeta(x)\}) = \int_M \mathbb{P}^t_{x,y}(A) p(t,x,y) \operatorname{vol}(\mathrm{d}y), \qquad A \in \mathscr{F}_s^{X(x)}$$

(which extends to $s = t$ because $\mathbb{P}^t_{x,y} \circ X_t(x)^{-1} = \delta_y$).
Since it is sufficient to verify this for $A = \{X_s \in B, s < \zeta(x)\}$, $B \in \mathscr{B}(M)$, the same computation as above yields

$$\mathbb{P}^t_{x,y}\{X_s \in B, s < \zeta(x)\} = \int_B \frac{p(s,x,\eta)p(t-s,\eta,y)}{p(t,x,y)} \operatorname{vol}(\mathrm{d}\eta)$$

and by integration with respect to $p(t,x,\,\cdot\,)\mathrm{d}\operatorname{vol}$ we conclude

$$\begin{aligned}\int_M \mathbb{P}^t_{x,y}(A) p(t,x,y) \operatorname{vol}(\mathrm{d}y) &= \int_M \int_B p(s,x,\eta)p(t-s,\eta,y) \operatorname{vol}(\mathrm{d}\eta) \operatorname{vol}(\mathrm{d}y) \\ &= \mathbb{P}(\{X_s(x) \in B, X_t(x) \in M, s < \zeta(x)\}) = \mathbb{P}(\{X_s(x) \in B, s < \zeta(x)\}).\end{aligned}$$

Finally, we obtain (2.32) by

$$\begin{aligned}d(P_t f)_x v &= \int_\Omega f \circ (X_t(x))(\omega) 1_{\{t<\zeta(\omega)\}} I_{\tau\wedge t}(\omega) \int_M \mathbb{P}^t_{x,y}(\mathrm{d}\omega) p(t,x,y) \operatorname{vol}(\mathrm{d}y) \\ &= \int_M f(y) p(t,x,y) \left[\int_\Omega I_{\tau\wedge t} \, \mathrm{d}\mathbb{P}^t_{x,y}\right] \operatorname{vol}(\mathrm{d}y).\end{aligned}$$

□

2.3. Possible choice of the finite energy process

We use this short paragraph to assure that the foregoing theorems do not treat the empty case, as it is not obvious whether such processes K_s exist or not. According to [Th 1], §4, we give a construction for the case of a regular domain, i.e. D is an open, relatively compact neighbourhood of x with smooth boundary.
This topic will be revisited in detail throughout Chapter 4.
The goal is to derive a nonnegative process k_s on $[0,t]$ which is adapted, bounded, has absolutely continuous paths, $k_0 = 1$, $k_{\tau\wedge t} = 0$ and $\left(\int_0^{\tau\wedge t} |\dot{k}_s|^2 \mathrm{d}s\right)^{1/2} \in L^p$ for some $p > 1$. Then $K_s = (k_s \wedge (1 - \frac{s}{t}))v$ can be used for the theorem.
Now let D have smooth boundary and choose some $f \in C^2(\bar{D})$ with $f|_{\partial D} = 0$ and $f|_D > 0$. By

$$T_s := \int_0^s \frac{\mathrm{d}r}{f(X_r(x))^2}\,, \quad s \le \tau, \qquad \text{and} \qquad \sigma_s = \inf\{r \ge 0 : T_r \ge s\}$$

we get two strictly increasing processes T_s and σ_s which are inverse to each other in the sense of $T_{\sigma_s} = s$, $s \le T_\tau$, and $\sigma_{T_s} = s$, $s \le \tau$ (recall that $\tau < \infty$ a.s. because of X_t being a non-degenerate diffusion process).
The time-changed process $X'_t(x) := X_{\sigma_t}(x)$ is a diffusion with generator $\widetilde{L} = f^2 L = f^2(\frac{1}{2}\Delta + V)$, which is verified by Itô's formula. Namely, for $g \in C^2(M)$ we have

$$g(X'_t(x)) - g(X'_0(x)) \stackrel{m}{=} \int_0^{\sigma_t} Lg(X_s(x))\mathrm{d}s = \int_0^{\sigma_t} \widetilde{L}g(X_s(x))\mathrm{d}T_s = \int_0^t \widetilde{L}g(X'_s(x))\mathrm{d}s.$$

To procede, we note a simple application of the chain rule.

Lemma 2.22. *For $f \in C^2(D)$ with $f > 0$ and any integer $m \in \mathbb{N}$ we have*

$$\Delta f^{-m} = m(m+1)\frac{|\operatorname{grad} f|^2}{f^{m+2}} - m\frac{\Delta f}{f^{m+1}}.$$

Proof. According to the composition formula (e.g. [H-Th], 7.155), for $\varphi :]0,\infty[\to]0,\infty[$, $x \mapsto x^{-m}$, we have

$$\begin{aligned}\Delta f^{-m} &= \Delta(\varphi \circ f) = \varphi_* \Delta f + \operatorname{trace}(f^* \nabla d\varphi) \\ &= -m f^{-m-1}\Delta f + m(m+1) f^{-m-2}\, |\operatorname{grad} f|^2 \\ &= -m\frac{\Delta f}{f^{m+1}} + m(m+1)\frac{|\operatorname{grad} f|^2}{f^{m+2}}\ .\end{aligned}$$

□

Now the most important observation concerning the time-changed diffusion is the following:

Lemma 2.23. *$X'_\cdot(x)$ is non-explosive, i.e. $\varrho := \inf\{t : X'_t(x) \in \partial D\} = \infty$ a.s.*

Proof. Let $X_0(x) = x \in D$ and fix some $n_0 \in \mathbb{N}$ with $f(x) \ge \frac{1}{n_0}$. We get an increasing sequence of stopping times $(\varrho_n)_{n \ge n_0}$ by $\varrho_n := \inf\{t : f(X'_t(x)) \le \frac{1}{n}\} \le \varrho$. So if we can

show that $\mathbb{P}\{\varrho_n < t\} \to 0$ for $n \to \infty$ and all $t > 0$, we also know that $\mathbb{P}\{\varrho < t\} = 0$ and the proof is finished.
But from $\Delta f^{-1} = -\Delta f/f^2 + 2|\operatorname{grad} f|^2/f^3$ on D (Lemma 2.22) we deduce

$$\widetilde{L} f^{-1} = f^2 \left(\frac{1}{2}\Delta + V\right) f^{-1} = -\frac{1}{2}\Delta f + |\operatorname{grad} f|^2/f + f^2 V(f^{-1}) \leq c f^{-1}$$

for some constant $c > 0$ because of f and all its derivatives being bounded on D as well as $f^2 V(f^{-1})$ since V is a first order derivative.
On the other hand, Itô's formula yields $\mathrm{d}(f^{-1} \circ X'_t) \overset{m}{=} \frac{1}{2}\widetilde{L} f^{-1}(X'_t)\mathrm{d}t$ and the omitted term is a true martingale if we stop time at ϱ_n for any $n \geq n_0$ since $f^{-1}(X'_{\bullet\wedge\varrho_n}(x)) \leq n$. So we find

$$\mathbb{E}\left[f^{-1}(X'_{t\wedge\varrho_n}(x))\right] - f^{-1}(x) = \frac{1}{2}\mathbb{E}\left[\int_0^{t\wedge\varrho_n} \widetilde{L} f^{-1}(X'_s(x))\mathrm{d}s\right] \leq \frac{c}{2}\int_0^t \mathbb{E}\left[f^{-1}(X'_{s\wedge\varrho_n}(x))\right]\mathrm{d}s.$$

By applying the Gronwall lemma to the continuous function $t \mapsto \mathbb{E}\left[f^{-1}(X'_{t\wedge\varrho_n}(x))\right]$ we end up with

$$n\mathbb{P}\{\varrho_n < t\} \leq \mathbb{E}\left[f^{-1}(X'_{t\wedge\varrho_n}(x))\right] \leq f^{-1}(x)\mathrm{e}^{\frac{c}{2}t}$$

for all $t > 0$ and $n \geq n_0$, which implies $\mathbb{P}\{\varrho_n < t\} \leq \frac{1}{n} f^{-1}(x)\mathrm{e}^{\frac{c}{2}t}$. $\square$

We note that because τ is the lifetime of X, T_τ has to be the lifetime of X' which in combination with the lemma means that $T_s = \int_0^s f^{-2}(X_r(x))\mathrm{d}r \to \infty$ for $s \nearrow \tau$ a.s. In particular, $T_{\sigma_t} = t$.
Now fix some $t > 0$ and choose a function $g \in C^1([0,t])$ with $g(0) = 1$, $g(t) = 0$. Then $k_s := g(T_s^{\sigma_t})$ is a real process on $[0,t]$ fulfilling the required properties, namely:

i) k is adapted, bounded, and has absolutely continuous paths.

ii) $k(0) = 1$, $k|_{[\sigma_t,t]} = 0$ and $\sigma_t \leq \tau \wedge t$.

iii) $\left(\int_0^{\tau\wedge t} |\dot{k}_s|^2 \mathrm{d}s\right)^{1/2} \in L^p$ for some $p > 1$.

Here this condition is satisfied e.g. for $p = 2$, because if $C_1 := \sup_{[0,t]} |\dot{g}|$ in the upper construction of k_s, then with $\mathrm{d}T_s = f^{-2}(X_s)\mathrm{d}s$ we find

$$\int_0^{\sigma_t} |\dot{k}_s|^2 \mathrm{d}s \leq C_1 \int_0^{\sigma_t} f^{-4}(X_s)\mathrm{d}s = C_1 \int_0^{\sigma_t} f^{-2}(X_s)\mathrm{d}T_s = C_1 \int_0^t f^{-2}(X'_s)\mathrm{d}s.$$

Again by Lemma 2.22 one has $\Delta f^{-2} = -2\Delta f/f^3 + 6|\nabla f|^2/f^4$ and this yields the existence of a $C_2 > 0$ such that $\widetilde{L} f^{-2} \leq C_2 f^{-2}$. As above, reasoning with Itô's formula and Gronwall for $\varrho'_n := \inf\{t \geq 0 : f^2(X'_t(x)) \leq \frac{1}{n}\}$ gives

$$\mathbb{E}[f^{-2}(X'_{s\wedge\varrho_n})] \leq f^{-2}(x)\mathrm{e}^{C_2 s/2}$$

for $n > f^{-2}(x)$. According to the lemma of Fatou we have

$$\mathbb{E}[f^{-2}(X'_s(x))] = \mathbb{E}[\lim_{n\to\infty} f^{-2}(X'_{s\wedge\varrho_n}(x))] \leq \liminf_{n\to\infty} \mathbb{E}[f^{-2}(X'_{s\wedge\varrho_n}(x))] \leq f^{-2}(x)\mathrm{e}^{C_2 s/2}$$

and the proof is finished.

Remark 2.24. For explicit estimates one may choose e.g.

$$f = \cos\left(\frac{\pi \operatorname{dist}(x,\,\cdot\,)}{2\operatorname{dist}(x,\partial D)}\right)$$

which is smooth on $D \setminus (\operatorname{cut}(x) \cup \{x\})$, where $\operatorname{cut}(x)$ denotes the *cut locus* of x.
We refer to Theorem 4.9, where this function is used in the proof.

2.4. Intrinsic gradient representation theorem

As we explained in the first paragraph of this chapter, the derivative process TX_s is only adapted to $\mathscr{F}^Z$, but not to $\mathscr{F}^{X(x)}$ ($x \in M$ a given and fixed initial point for the diffusion). Consequently, the foregoing results (which all contain terms of $TX_s \equiv X_{s*}$) do not only depend on the geometrical situation, i.e. the given Riemannian manifold (M, g) and the elliptic generator L, but also on the driving Brownian motion Z, or equivalently, on a given isometric embedding $M \to \mathbb{R}^m$. For this reason, we may call the preceding formulae 'non-intrinsic'. But with the results of Proposition 2.10 or, equivalently, Theorem 2.12 we can easily translate the first order derivative formulae to the intrinsic case by replacing X_{s*} by W_s given in (2.8).
The subsequent results hold if ∇ denotes the Le Jan-Watanabe connection. We once again point out that it coincides with the Levi-Civita connection if (M, g) and A form a gradient Brownian system.

Theorem 2.25. *Let $(K_s)_{s\in[0,t]}$ be an $\left(\mathscr{F}_s^{X(x)}\right)$-adapted process with sample paths in the Cameron-Martin space $\mathbb{H}_0([0,t], T_xM)$. Then $(n_s)_{0\le s\le \sigma\wedge t}$ given by*

$$n_s := (dF_s)_{X_s(x)} W_s K_s - F_s(X_s(x)) \int_0^s \langle W_r \dot{K}_r, A(X_r(x))\mathrm{d}Z_r\rangle \tag{2.34}$$

is a $\left(\mathscr{F}_s^{X(x)}\right)$-adapted local martingale, where $\sigma < \zeta(x)$ is a stopping time.

Proof. Because both W_s and $A(X_s)\mathrm{d}Z_s = \mathrm{d}X_s - A_0(X_s)\mathrm{d}s$ are $(\mathscr{F}^{X(x)})$-adapted, so is the integral in formula (2.34) and therefore (n_s) itself. By Theorem 2.12 $(dF_s)_{X_s(x)}W_s$ is a local martingale taking values in T_x^*M and we could repeat the integration by parts proof of paragraph 2.2.1 with this process instead of m_s there.
Alternatively, we may take the local martingale n_s given by formula (2.12) and derive (2.34) by filtering out redundant noise, i.e. taking conditional expectation $\tilde{n}_s := \mathbb{E}[n_s|\mathscr{F}_s^{X(x)}]$ according to Proposition 2.10.
The only difficulty then is to verify that $\mathbb{E}[T_xX_r|\mathscr{F}_s^{X(x)}] = \mathbb{E}[T_xX_r|\mathscr{F}_r^{X(x)}] = W_r$ for arbitrary $0 \le r \le s$, since one has to take the conditional expectation under the integral (and all the other factors there are adapted to the smaller filtration).
But this is clear if one goes back to the (heuristic) argument subsequent to Proposition 2.10 where we explained that filtering out noise from the derivative process means to cancel the part of Z^0 in the orthogonal decomposition of Z with respect to the kernel of A along the paths of T_xX_s. ☐

We mention the existence of several possibilities to get the intrinsic local martingale, because in the second derivative case, some of them may work and others fail.

Remark 2.26. Although we saw that $A(X_s(x))\mathrm{d}Z_s$ is adapted with respect to the smaller filtration $\big(\mathscr{F}_s^{X(x)}\big)$, one tends to replace it in intrinsic formulae by terms of stochastic parallel translation along paths of $X_s(x)$ and the related n-dimensional Brownian motion B_s on T_xM from (2.6), which was given by

$$B_s := \int_0^s //_{0,r}^{-1} A(X_r(x))\mathrm{d}Z_r.$$

So we can reread the definition of B_s the following way:

$$A(X_s(x))\mathrm{d}Z_s = //_{0,s}\mathrm{d}B_s. \tag{2.35}$$

Hence, the upper local martingale can also be written as

$$n_s := (dF_s)_{X_s(x)} W_s K_s - F_s(X_s(x)) \int_0^s \langle W_r \dot{K}_r, //_{0,r}\mathrm{d}B_r\rangle. \tag{2.36}$$

So the main intrinsic result for first order derivatives of heat semigroups can be stated as follows.

Theorem 2.27 (Intrinsic representation formula for the gradient).
Again let $\zeta(\,\cdot\,)$ denote the lifetime of $X_t(\,\cdot\,)$, fix $x \in M$ and let τ be the first exit time of $X(x)$ from a relatively compact neighbourhood D of x. We assume $t > 0$, $F : [0,t]\times M \to \mathbb{R}$ to be space-time-harmonic and $v \in T_xM$.
Then for every bounded adapted process $(K_s)_{s\in[0,t]}$ with sample paths in $\mathbb{H}([0,t], T_xM)$, $K_0 = v$, $K_s|_{s\in[\tau\wedge t,t]} = 0$ and $(\int_0^t \|\dot{K}_s\|^2\mathrm{d}s)^{1/2} \in L^{1+\alpha}(\mathbb{P})$ for some $\alpha > 0$, we have

$$\langle \operatorname{grad}_x F_0, v\rangle \equiv (dF_0)_x v = -\mathbb{E}\left[(F_{\tau\wedge t}\circ X_{\tau\wedge t}(x)) \int_0^{\tau\wedge t} \langle W_s \dot{K}_s, //_{0,s}\mathrm{d}B_s\rangle\right], \tag{2.37}$$

which in the heat semigroup case for bounded measurable $f : M \to \mathbb{R}$ can be modified to

$$\langle \operatorname{grad}_x P_t f, v\rangle \equiv d(P_t f)_x v = -\mathbb{E}\left[(f\circ X_t(x)) 1_{\{t<\zeta(x)\}} \int_0^t \langle W_s \dot{K}_s, //_{0,s}\mathrm{d}B_s\rangle\right]. \tag{2.38}$$

Proof. The proof exactly is the same as for Theorem 2.19 now using the intrinsic (local) martingale. One should notice that taking the conditional expectation is contractive, i.e. on $\mathrm{Lin}(T_xM, T_{X_s(x)}M)$ we have

$$\|W_s\| = \|//_{0,s}\mathbb{E}[//_{0,s}^{-1} T_x X_s | \mathscr{F}_s^{X(x)}]\| \le \|\mathbb{E}[//_{0,s}^{-1} T_x X_s | \mathscr{F}_s^{X(x)}]\| \le \|//_{0,s}^{-1} T_x X_s\| \le \|T_x X_s\|$$

since the stochastic parallel translation is isometric. □

We finally rewrite Corollary 2.21 in intrinsic terms.

Corollary 2.28. *Under the assumptions of Theorem 2.27 the Bismut formula for the gradient of the logarithmic heat kernel reads*

$$\begin{aligned}\langle \operatorname{grad}_x(\log p(t,\,\cdot\,,y)), v\rangle &\equiv d(\log p(t,\,\cdot\,,y))_x v \\ &= \frac{d(p(t,\,\cdot\,,y))_x v}{p(t,x,y)} = \mathbb{E}\left[I_{\tau\wedge t} | X_t(x) = y\right],\end{aligned} \tag{2.39}$$

where $I_{\tau\wedge t}$ is given by

$$I_r := -\int_0^r \langle W_s \dot{K}_s, //_{0,s}\mathrm{d}B_s\rangle. \tag{2.40}$$

Chapter 3

Stochastic representation formulae for the Hessian

In this paragraph we extend the methods described in the preceding chapter in order to compute second order derivatives. More exactly, we obtain stochastic representation formulae for the Hessian

$$\operatorname{Hess}_x F_0(v,w) \equiv (\nabla dF_0)(x)(v,w), \quad v,w \in T_xM,$$

where F is space-time-harmonic just as in the previous chapters. Of course this includes the heat semigroup case as well as harmonic functions.
One should recall that for $f \in C^2(M)$ the Hessian

$$(\nabla df)(x)(\,\cdot\,,\,\cdot\,) \equiv \nabla\,.\,(df)_x(\,\cdot\,)$$

is a smooth bilinear form on $T_xM \times T_xM$ which is symmetric for torsion-free connections as the Levi-Civita and the Le Jan-Watanabe connections are.
For brevity, we omit the base point x on M at which the Hessian is taken and write $\nabla dF_0(v,w)$ if it is clear to which tangent space the arguments belong to.
Requiring rather strong conditions, such formulae were first introduced by Elworthy and Li, cf. [E-L 1], Thm. 3.1, and proven by a differentiation under the expectation argument. For this reason, their result was only stated for the heat semigroup applied to C^2-functions on a compact manifold.
With our methods we find several ways to prove that some stochastic process in terms of X_s and its (spatial) first and second order derivatives is a local martingale, which is suitable to derive non-intrinsic expressions for the Hessian of P_tf. These proofs will be carried out in detail because they are new and later on we have to decide which methods can be transferred to intrinsic formulas and which fail to do so.

The main difference to the first order formulae is that we will mostly use two finite energy processes K and L instead of one, which provides another degree of freedom in the results. However, to have concise formulas and to be able to filter out noise, it turns out that these processes should be independent in the sense that they vary on disjoint (stochastic) time intervals, i.e. one of them decays from $v \in T_xM$ to 0 within time $[0,\sigma]$ and the other one remains constantly at $w \in T_xM$ until σ and declines to 0 in a further interval $]\sigma,\tau]$.

3.1. Non-intrinsic calculations

The most important observation concerning non-intrinsic second order derivatives is the following one:

Theorem 3.1. *If (K_s) and (L_s) are two $(\mathscr{F}_s^Z)$-adapted processes on a finite time horizon $[0,t]$ with sample paths in $\mathbb{H}([0,t],T_xM)$ for some fixed $x \in M$ and deterministic initial values $K_0 = v$, $L_0 = w$, then the following process is a local martingale on $[0,t]$:*

$$\begin{aligned} n_s := {} & \nabla dF_s(X_{s*}K_s, X_{s*}L_s) + (dF_s)_{X_s(x)} \nabla TX_s(K_s, L_s) \\ & - (dF_s)_{X_s(x)} X_{s*} \int_0^s X_{r*}^{-1} \Big[\nabla TX_r(K_r, \dot{L}_r) + \nabla TX_r(L_r, \dot{K}_r) \Big] \, \mathrm{d}r \\ & + (dF_s)_{X_s(x)} X_{s*} \int_0^s X_{r*}^{-1} \Big[\nabla A(X_{r*}K_r) A(X_r(x))^* X_{r*} \dot{L}_r \\ & \qquad\qquad + \nabla A(X_{r*}L_r) A(X_r(x))^* X_{r*} \dot{K}_r \Big] \, \mathrm{d}r \\ & - (dF_s)_{X_s(x)} X_{s*} K_s \int_0^s \langle X_{r*}\dot{L}_r, A(X_r(x)) \mathrm{d}Z_r \rangle \\ & - (dF_s)_{X_s(x)} X_{s*} L_s \int_0^s \langle X_{r*}\dot{K}_r, A(X_r(x)) \mathrm{d}Z_r \rangle \\ & + F_s(X_s(x)) \bigg(\int_0^s \langle X_{r*}\dot{K}_r, A(X_r(x)) \mathrm{d}Z_r \rangle \int_0^s \langle X_{r*}\dot{L}_r, A(X_r(x)) \mathrm{d}Z_r \rangle \\ & \qquad\qquad - \int_0^s \langle X_{r*}\dot{K}_r, X_{r*}\dot{L}_r \rangle \mathrm{d}r \bigg). \end{aligned} \tag{3.1}$$

On paths where $\zeta(x) \le t$, we assume that K and L equal 0 from a stopping time $\tau < \zeta(x)$ up to t. Then all terms are well-defined because of their bilinear dependence on (K, L).

Before we give two proofs of this, we explain the expressions containing covariant derivatives which appear for the first time here (in fact, we already have considered ∇A in the defining equation of the derivative process):
∇TX_s stands for the covariant derivative of the derivative process and therefore it is (a.s.) the result of (covariantly) differentiating twice the diffeomorphism $X_s(\,\cdot\,,\omega) : M_s(\omega) \to M$, thus it is a bilinear symmetric map. In fact, TX_s is a section in $\Gamma(T^*M_s \otimes X_s^*TM) \equiv \mathrm{Hom}(TM_s, X_s^*TM)$, where as in Chapter 2 we have to take in account that the diffusion's possibly finite lifetime demands restriction to the open subset M_s of M (depending on ω). For this reason we have

$$\begin{aligned} \nabla TX_s \in \Gamma(T^*M_s \otimes T^*M_s \otimes X_s^*TM) &\equiv \mathrm{Hom}(TM_s \otimes TM_s, X_s^*TM) \\ &\equiv \mathrm{Bil}(TM_s, TM_s; X_s^*TM). \end{aligned}$$

In each case ∇ denotes the connection on the corresponding tensor bundle induced by the Levi-Civita connection (or the Le Jan-Watanabe connection in the corresponding situation) on TM, which is defined via the Leibniz rule. Some of the factors here contain the tangent or cotangent bundle over the open submanifold M_s of M (ω by ω), but this raises no additional problems since connections are local.

Concerning ∇A at last, we have

$$\nabla A \in \Gamma(\mathbb{R}^m \otimes T^*M \otimes TM) \equiv \mathrm{Hom}(\mathbb{R}^m, \mathrm{Hom}(TM, TM)).$$

because $A \in \Gamma(\mathbb{R}^m \otimes TM) \equiv \mathrm{Hom}(\mathbb{R}^m, TM)$.
According to this equivalence we tend to write $\nabla A(v)z$ for $v \in T_xM$ and $z \in \mathbb{R}^m$ rather than $\nabla A(z, v)$. In equation (3.1) we make use of the composition

$$\nabla A(X_{s*}(\,\cdot\,)) \in \mathrm{Hom}(\mathbb{R}^m, \mathrm{Hom}(TM_s, X^*TM)).$$

For the proof of Theorem 3.1 we start out by the perturbation argument this time because in this case it is more or less obvious how to proceed.

3.1.1. Perturbation-Girsanov type proof

On principle, to derive second order derivatives we on have two possibilities: Either take the initial martingale $F_s(X_s(x))$, modify the driving diffusion $X_s(x)$ by a two-parameter perturbation and differentiate with respect to both at 0, or start with the first derivative local martingale (2.12), perturb $X_s(x)$ in here by just one parameter and differentiate.
In the end, both methods are of comparable effort, the first one is straightforward and completely symmetric in K_s and L_s but all terms contained herein are related to two parameters ε and δ, the second one involves some more complicated partial differentiations of parameter-dependent homomorphisms on the tangent bundle and includes some additional integration by parts arguments at the end.

First perturbation type proof of Theorem 3.1. We introduce two real parameters ε and δ and start out with the perturbed equation

$$\mathrm{d}X_s^{\varepsilon,\delta} = A(X_s^{\varepsilon,\delta}) * \mathrm{d}Z_s^{\varepsilon,\delta} + A_0(X_s^{\varepsilon,\delta})\mathrm{d}s \tag{3.2}$$

where for two $\mathbb{R}^m$-valued adapted processes k_s and ℓ_s from $L^2_{\mathrm{loc}}(\mathrm{d}s)$ the driving process $Z_s^{\varepsilon,\delta}$ shall be given by

$$\mathrm{d}Z_s^{\varepsilon,\delta} = \mathrm{d}Z_s + (\varepsilon k_s + \delta \ell_s)\mathrm{d}s.$$

(as in the gradient case we can assume that k_s, $\ell_s \in (\ker A(X_s(x)))^\perp$).
For this purpose, we define $H_s^{\varepsilon,\delta}(x)$ to be the solution of the pathwise equation

$$\begin{aligned} \mathrm{d}H_s^{\varepsilon,\delta} &= X_{s*}^{-1} A(X_s \circ H_s^{\varepsilon,\delta})(\varepsilon k_s + \delta \ell_s)\mathrm{d}s, \\ H_s^{0,0} &= \mathrm{id}_M \end{aligned} \tag{3.3}$$

and again let the disturbed solution factorize over the unperturbed one by

$$X_s^{\varepsilon,\delta}(x) := X_s \circ H_s^{\varepsilon,\delta}(x), \qquad x \in M. \tag{3.4}$$

As in the first derivative case, the generalized Itô formula yields the upper representation of $Z_s^{\varepsilon,\delta}(x)$, cf. (2.18). It should be noticed that we postpone the choice of the initial value $H_0^{\varepsilon,\delta}(x)$ and return to this point later.

The corresponding Girsanov correction factor this time is given by the exponential (local) martingale

$$\begin{aligned} G_s^{\varepsilon,\delta} &= \mathscr{E}\Big(-\int_0^s \langle \varepsilon k_r + \delta \ell_r, \mathrm{d}Z_r\rangle\Big) \\ &= \exp\left(-\int_0^s \langle \varepsilon k_r + \delta \ell_r, \mathrm{d}Z_r\rangle - \frac{1}{2}\int_0^s \|\varepsilon k_r + \delta \ell_r\|^2 \,\mathrm{d}r\right). \end{aligned} \tag{3.5}$$

So we have the C^2-family of real local martingales $(\varepsilon,\delta) \mapsto (F_s \circ X_s^{\varepsilon,\delta}(x))G_s^{\varepsilon,\delta}$ of which we compute partial derivatives with respect to both parameters. To put emphasis on the fact that we end up in $T\mathbb{R}$ after the first differentiation we oppress the identification $T\mathbb{R} \equiv \mathbb{R}$ and write the covariant symbol for the second one.
Partial differentiation by the product rule leads to the local martingale

$$\begin{aligned} n_s :=& \frac{\nabla}{\partial\delta}\Big|_{\delta=0} \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} \Big(\Big(F_s \circ X_s^{\varepsilon,\delta}(x)\Big) G_s^{\varepsilon,\delta}\Big) \\ =& \frac{\nabla}{\partial\delta}\Big|_{\delta=0} \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} \Big(F_s \circ X_s^{\varepsilon,\delta}(x)\Big) G_s^{0,0} + \frac{\partial}{\partial\delta}\Big|_{\delta=0} \Big(F_s \circ X_s^{0,\delta}(x)\Big) \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} G_s^{\varepsilon,0} \\ &+ \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} \big(F_s \circ X_s^{\varepsilon,0}(x)\big) \frac{\partial}{\partial\delta}\Big|_{\delta=0} G_s^{0,\delta} + F_s(X_s(x)) \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} \frac{\partial}{\partial\delta}\Big|_{\delta=0} G_s^{\varepsilon,\delta}. \end{aligned} \tag{3.6}$$

We mention that if we are given a torsion free connection such as the Levi-Civita connection here, the Schwarz lemma also holds for second (covariant) partial derivatives of manifold-valued functions, cf. [Kl], Prop. 1.5.8. Thus we do not need to take care of the order of taking partial derivatives.
The differentiation of the stochastic exponential in (3.5) leads to

$$\begin{aligned} G_s^{0,0} = 1, \quad \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} G_s^{\varepsilon,0} = -\int_0^s \langle k_r, \mathrm{d}Z_r\rangle, \quad \frac{\partial}{\partial\delta}\Big|_{\delta=0} G_s^{0,\delta} = -\int_0^s \langle \ell_r, \mathrm{d}Z_r\rangle, \\ \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} \frac{\partial}{\partial\delta}\Big|_{\delta=0} G_s^{\varepsilon,\delta} = \int_0^s \langle k_r, \mathrm{d}Z_r\rangle \int_0^s \langle \ell_r, \mathrm{d}Z_r\rangle - \int_0^s \langle k_r, \ell_r\rangle \mathrm{d}r \end{aligned}$$

and for the second and third term on the right hand side of (3.6) we find just as in the first order case that

$$\frac{\partial}{\partial\delta}\Big|_{\delta=0} \Big(F_s \circ X_s^{0,\delta}(x)\Big) = (dF_s)_{X_s(x)} X_{s*} \frac{\partial}{\partial\delta}\Big|_{\delta=0} H_s^{0,\delta}(x)$$

and

$$\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} \big(F_s \circ X_s^{\varepsilon,0}(x)\big) = (dF_s)_{X_s(x)} X_{s*} \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} H_s^{\varepsilon,0}(x).$$

So the main work is to compute

$$\frac{\nabla}{\partial\delta}\Big|_{\delta=0} \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} \Big(F_s(X_s \circ H_s^{\varepsilon,\delta}(x))\Big) = \frac{\nabla}{\partial\delta}\Big|_{\delta=0} \left((dF_s)_{X_s^{0,\delta}} T_{H_s^{0,\delta}(x)} X_s \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} H_s^{\varepsilon,\delta}(x)\right).$$

Herein, on the right hand side we have three terms to be differentiated with respect to δ by the Leibniz rule. For this reason the last term equals

$$\nabla dF_s \left(X_{s*} \frac{\partial}{\partial\delta}\Big|_{\delta=0} H_s^{0,\delta}(x), X_{s*} \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} H_s^{\varepsilon,0}(x)\right)$$

$$+ (dF_s)_{X_s(x)} \nabla T X_s \left(\frac{\partial}{\partial \delta}\Big|_{\delta=0} H_s^{0,\delta}(x), \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} H_s^{\varepsilon,0}(x) \right)$$
$$+ (dF_s)_{X_s(x)} X_{s*} \frac{\nabla}{\partial \delta}\Big|_{\delta=0} \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} H_s^{\varepsilon,\delta}(x).$$

Now we choose initial values

$$\frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} H_0^{\varepsilon,0}(x) = v, \qquad \frac{\partial}{\partial \delta}\Big|_{\delta=0} H_0^{0,\delta}(x) = w, \qquad \frac{\nabla}{\partial \delta}\Big|_{\delta=0} \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} H_0^{\varepsilon,\delta}(x) = 0, \tag{3.7}$$

$v, w \in T_xM$, which is easily achieved by a proper choice of the C^∞-mapping $(\varepsilon, \delta) \mapsto H_0^{\varepsilon,\delta}(x)$ at $(0,0)$.
So by reasoning as in (2.21) (because $H_s^{\varepsilon,0}(x)$ and $H_s^{0,\delta}(x)$ contain just one perturbation) we obtain and define

$$K_s := \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} H_s^{\varepsilon,0}(x) = v + \int_0^s X_{r*}^{-1} A(X_r(x)) k_r \mathrm{d}r \tag{3.8}$$

and

$$L_s := \frac{\partial}{\partial \delta}\Big|_{\delta=0} H_s^{0,\delta}(x) = w + \int_0^s X_{r*}^{-1} A(X_r(x)) \ell_r \mathrm{d}r. \tag{3.9}$$

Differentiation provides that k_s and ℓ_s now can be replaced by the following expressions (since $k_s = A(X_s(x))^* A(X_s(x)) k_s$ and the same holds for ℓ_s):

$$k_s = A(X_s(x))^* X_{s*} \dot{K}_s, \qquad \ell_s = A(X_s(x))^* X_{s*} \dot{L}_s. \tag{3.10}$$

Hence, $\langle k_s, \ell_s \rangle = \langle X_{s*} \dot{K}_s, X_{s*} \dot{L}_s \rangle$.
Now we can put all these results and definitions in equation (3.6) to derive the preliminary result

$$\begin{aligned} n_s = {} & \nabla dF_s(X_{s*}K_s, X_{s*}L_s) + (dF_s)_{X_s(x)} \nabla T X_s(K_s, L_s) \\ & + (dF_s)_{X_s(x)} X_{s*} \frac{\nabla}{\partial \delta}\Big|_{\delta=0} \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} H_s^{\varepsilon,\delta}(x) \\ & - (dF_s)_{X_s(x)} X_{s*} K_s \int_0^s \langle X_{r*} \dot{L}_r, A(X_r(x)) \mathrm{d}Z_r \rangle \\ & - (dF_s)_{X_s(x)} X_{s*} L_s \int_0^s \langle X_{r*} \dot{K}_r, A(X_r(x)) \mathrm{d}Z_r \rangle \\ & + F_s(X_s(x)) \bigg(\int_0^s \langle X_{r*} \dot{K}_r, A(X_r(x)) \mathrm{d}Z_r \rangle \int_0^s \langle X_{r*} \dot{L}_r, A(X_r(x)) \mathrm{d}Z_r \rangle \\ & \qquad - \int_0^s \langle X_{r*} \dot{K}_r, X_{r*} \dot{L}_r \rangle \mathrm{d}r \bigg). \end{aligned} \tag{3.11}$$

So the aim for rest of the proof is to compute $\frac{\nabla}{\partial \delta}\big|_{\delta=0} \frac{\partial}{\partial \varepsilon}\big|_{\varepsilon=0} H_s^{\varepsilon,\delta}(x)$.
We again assume that we are allowed to integrate the (with respect to ε and δ) formally differentiated version of equation (3.3) defining $H_s^{\varepsilon,\delta}(x)$ to derive this term and postpone the justification to Remark 3.2. This yields

$$\frac{\nabla}{\partial \delta}\Big|_{\delta=0} \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} H_s^{\varepsilon,\delta}(x)$$

$$
\begin{aligned}
&= \frac{\nabla}{\partial\delta}\Big|_{\delta=0} \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} H_0^{\varepsilon,\delta}(x) \\
&\quad + \int_0^s \frac{\nabla}{\partial\delta}\Big|_{\delta=0} \frac{\nabla}{\partial\varepsilon}\Big|_{\varepsilon=0} \left(T_{X_r^{\varepsilon,\delta}(x)} X_r^{-1} A(X_r \circ H_r^{\varepsilon,\delta}(x))(\varepsilon k_r + \delta \ell_r)\right) \mathrm{d}r \\
&= \int_0^s \frac{\nabla}{\partial\varepsilon}\Big|_{\varepsilon=0} \frac{\nabla}{\partial\delta}\Big|_{\delta=0} \left[\delta\left(T_{X_r^{\varepsilon,\delta}(x)} X_r^{-1} A(X_r \circ H_r^{\varepsilon,\delta}(x))\ell_r\right)\right] \mathrm{d}r \\
&\quad + \int_0^s \frac{\nabla}{\partial\delta}\Big|_{\delta=0} \frac{\nabla}{\partial\varepsilon}\Big|_{\varepsilon=0} \left[\varepsilon\left(T_{X_r^{\varepsilon,\delta}(x)} X_r^{-1} A(X_r \circ H_r^{\varepsilon,\delta}(x))k_r\right)\right] \mathrm{d}r \\
&= \int_0^s \frac{\nabla}{\partial\varepsilon}\Big|_{\varepsilon=0} \left(T_{X_r^{\varepsilon,0}(x)} X_r^{-1} A(X_r \circ H_r^{\varepsilon,0}(x))\right) \ell_r \mathrm{d}r \\
&\quad + \int_0^s \frac{\nabla}{\partial\delta}\Big|_{\delta=0} \left(T_{X_r^{0,\delta}(x)} X_r^{-1} A(X_r \circ H_r^{0,\delta}(x))\right) k_r \mathrm{d}r.
\end{aligned}
$$

In the first integral we have

$$
\begin{aligned}
&\frac{\nabla}{\partial\varepsilon}\Big|_{\varepsilon=0} \left(T_{X_r^{\varepsilon,0}(x)} X_r^{-1} A(X_r \circ H_r^{\varepsilon,0}(x))\right) \ell_r \\
&= \nabla T X_r^{-1}\left(X_{r*} \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} H_r^{\varepsilon,0}(x), A(X_r(x))\ell_r\right) + X_{r*}^{-1} \nabla A \left(X_{r*} \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} H_r^{\varepsilon,0}(x)\right) \ell_r \\
&= -X_{r*}^{-1} \nabla T X_r(K_r, \dot{L}_r) + X_{r*}^{-1} \nabla A(X_{r*}K_r) A(X_r(x))^* X_{r*}\dot{L}_r
\end{aligned}
$$

because of $\dot{L}_r = X_{r*}^{-1} A(X_r(x))\ell_r$. We also used that by covariantly differentiating the identity $\mathrm{id}_{T_xM} = T_{X_r(x)} X_r^{-1} T_x X_r$ according to the product rule we obtain

$$\nabla T X_r^{-1}(X_{r*}\,\cdot\,, X_{r*}\,\cdot\,) = -X_{r*}^{-1} \nabla T X_r(\,\cdot\,,\,\cdot\,).$$

The analogous calculation for the second term gives

$$
\begin{aligned}
&\frac{\nabla}{\partial\delta}\Big|_{\delta=0} \left(T_{X_r^{0,\delta}(x)} X_r^{-1} A(X_r \circ H_r^{0,\delta}(x))\right) k_r \\
&= -X_{r*}^{-1} \nabla T X_r(\dot{K}_r, L_r) + X_{r*}^{-1} \nabla A(X_{r*}L_r) A(X_r(x))^* X_{r*}\dot{K}_r.
\end{aligned}
$$

So we can put the result

$$
\begin{aligned}
&\frac{\nabla}{\partial\delta}\Big|_{\delta=0} \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} H_s^{\varepsilon,\delta}(x) \\
&\quad = \int_0^s \left[-X_{r*}^{-1} \nabla T X_r(K_r, \dot{L}_r) + X_{r*}^{-1} \nabla A(X_{r*}K_r) A(X_r(x))^* X_{r*}\dot{L}_r\right] \mathrm{d}r \qquad (3.12) \\
&\quad\quad + \int_0^s \left[-X_{r*}^{-1} \nabla T X_r(\dot{K}_r, L_r) + X_{r*}^{-1} \nabla A(X_{r*}L_r) A(X_r(x))^* X_{r*}\dot{K}_r\right] \mathrm{d}r
\end{aligned}
$$

into equation (3.11) to end up with (3.1). □

Remark 3.2 *(On the formal differentiation of the defining equation for* $H_s^{\varepsilon,\delta}$*).*
From (3.3) we have

$$f \circ H_s^{\varepsilon,\delta}(x) = f \circ H_0^{\varepsilon,\delta}(x) + \int_0^s (df)_{H_r^{\varepsilon,\delta}(x)} X_{r*}^{-1} A(X_r \circ H_r^{\varepsilon,\delta}(x))(\varepsilon k_r + \delta \ell_r)\mathrm{d}r$$

for all $f \in C_c^\infty(M)$. By partial differentiation with respect to ε and δ at 0 and taking derivatives under the integral sign we obtain

$$\begin{aligned}
&\nabla df\left(\frac{\partial}{\partial\delta}\Big|_{\delta=0} H_s^{0,\delta}(x), \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} H_s^{\varepsilon,0}(x)\right) + (df)_x \frac{\nabla}{\partial\delta}\Big|_{\delta=0}\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} H_s^{\varepsilon,\delta}(x)\\
&= \nabla df\left(\frac{\partial}{\partial\delta}\Big|_{\delta=0} H_0^{0,\delta}(x), \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} H_0^{\varepsilon,0}(x)\right) + (df)_x \frac{\nabla}{\partial\delta}\Big|_{\delta=0}\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} H_0^{\varepsilon,\delta}(x)\\
&\quad + \int_0^s \frac{\nabla}{\partial\delta}\Big|_{\delta=0}\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0}\left[(df)_{H_r^{\varepsilon,\delta}(x)} X_{r*}^{-1} A(X_r \circ H_r^{\varepsilon,\delta}(x))(\varepsilon k_r + \delta \ell_r)\mathrm{d}r\right].
\end{aligned}$$

Using (3.7) and the definitions of K_s and L_s from above we can rewrite this as

$$\begin{aligned}
&\nabla df(K_s, L_s) + (df)_x \frac{\nabla}{\partial\delta}\Big|_{\delta=0}\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} H_s^{\varepsilon,0}(x)\\
&\quad= \nabla df(v,w) + \int_0^s \frac{\nabla}{\partial\delta}\Big|_{\delta=0}\left[(df)_{H_r^{0,\delta}(x)} X_{r*}^{-1} A(X_r \circ H_r^{0,\delta}(x))k_r\right]\mathrm{d}r\\
&\qquad\qquad + \int_0^s \frac{\nabla}{\partial\varepsilon}\Big|_{\varepsilon=0}\left[(df)_{H_r^{\varepsilon,0}(x)} X_{r*}^{-1} A(X_r \circ H_r^{\varepsilon,0}(x))\ell_r\right]\mathrm{d}r\\
&\quad= \nabla df(v,w) + \int_0^s \nabla df\left(\frac{\partial}{\partial\delta}\Big|_{\delta=0} H_r^{0,\delta}(x), X_{r*}^{-1}A(X_r(x))k_r\right)\mathrm{d}r\\
&\qquad\qquad + \int_0^s \nabla df\left(\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} H_r^{\varepsilon,0}(x), X_{r*}^{-1}A(X_r(x))\ell_r\right)\mathrm{d}r\\
&\qquad\qquad + \int_0^s (df)_x \frac{\nabla}{\partial\delta}\Big|_{\delta=0}\left(X_{r*}^{-1}A(X_r\circ H_r^{0,\delta}(x))k_r\right)\mathrm{d}r\\
&\qquad\qquad + \int_0^s (df)_x \frac{\nabla}{\partial\varepsilon}\Big|_{\varepsilon=0}\left(X_{r*}^{-1}A(X_r\circ H_r^{\varepsilon,0}(x))\ell_r\right)\mathrm{d}r\\
&\quad= \nabla df(v,w) + \int_0^s \nabla df(\dot K_r, L_r)\mathrm{d}r + \int_0^s \nabla df(K_r, \dot L_r)\mathrm{d}r\\
&\qquad\qquad + \int_0^s (df)_x \frac{\nabla}{\partial\delta}\Big|_{\delta=0}\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0}\left(X_{r*}^{-1}A(X_r\circ H_r^{\varepsilon,\delta}(x))(\varepsilon k_r + \delta\ell_r)\right)\mathrm{d}r.
\end{aligned}$$

Since K_s and L_s are processes with paths of bounded variation on the finite time horizon $[0,t]$ we can apply integration by parts with respect to the bilinear symmetric form ∇df on T_xM to derive

$$\begin{aligned}
\nabla df(K_s, L_s) &= \nabla df(K_0, L_0) + \int_0^s \nabla df(\mathrm{d}K_r, L_r) + \int_0^s \nabla df(K_r, \mathrm{d}L_r)\\
&= \nabla df(v,w) + \int_0^s \nabla df(\dot K_r, L_r)\mathrm{d}r + \int_0^s \nabla df(K_r, \dot L_r)\mathrm{d}r.
\end{aligned}$$

Inserting this into the upper equation and using a proper variety of smooth functions f we get

$$\frac{\nabla}{\partial\delta}\Big|_{\delta=0}\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} H_s^{\varepsilon,\delta}(x) = \int_0^s \frac{\nabla}{\partial\delta}\Big|_{\delta=0}\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0}\left(X_{r*}^{-1}A(X_r\circ H_r^{\varepsilon,\delta}(x))(\varepsilon k_r + \delta\ell_r)\right)\mathrm{d}r$$

which is the integrated version of the formally differentiated equation (3.3).

Remark 3.3 *(Continuation of Remark 2.17).*
Instead of demanding that $H_0^{\varepsilon,\delta}(x)$ fulfills the conditions in (3.3), we could have set $H_0^{\varepsilon,\delta} := \mathrm{id}_M$ in the proof above, thus $\frac{\partial}{\partial\varepsilon}\big|_{\varepsilon=0} H_0^{\varepsilon,0}(x) = \frac{\partial}{\partial\delta}\big|_{\delta=0} H_0^{0,\delta}(x) = \frac{\nabla}{\partial\delta}\big|_{\delta=0} \frac{\partial}{\partial\varepsilon}\big|_{\varepsilon=0} H_0^{\varepsilon,\delta}(x) = 0$.
To compensate this simplification, however, there need some extra computations to be done afterwards.
Briefly described, we then have the upper result for the two processes $\widetilde{K}_s := K_s - v$ and $\widetilde{L}_s := L_s - w$ starting in 0 and write $\tilde{n}_s$ for this new local martingale. Observe that $\dot{K}_s = \dot{\widetilde{K}}_s$ and $\dot{L}_s = \dot{\widetilde{L}}_s$.
Consequently, we obtain the local martingale property of n_s by adding some local martingales to $\tilde{n}_s$, as there are

$$\nabla_v \big((dF_s)_{X_s(x)} X_{s*} w\big) = \nabla dF_s(X_{s*}v, X_{s*}w) + (dF_s)_{X_s(x)} \nabla T X_s(v,w)$$

and the covariant differentials of the first order local martingales

$$\nabla_w \left[(dF_s)_{X_s(x)} X_{s*} K_s - F_s(X_s(x)) \int_0^s \langle X_{r*} \dot{K}_r, A(X_r(x))\mathrm{d}Z_r\rangle\right]$$

and

$$\nabla_v \left[(dF_s)_{X_s(x)} X_{s*} L_s - F_s(X_s(x)) \int_0^s \langle X_{r*} \dot{L}_r, A(X_r(x))\mathrm{d}Z_r\rangle\right].$$

By the methods of the preceding proof it is clear how to carry this out.

The other possibility to derive the general second order local martingale by perturbation methods is the following.

Second perturbation type proof of Theorem 3.1. Instead of a two-parameter perturbation of the basic local martingale $F_s \circ X_s(x)$, we introduce a one-parameter perturbation of the local martingale (2.12) that we obtained in the gradient case. Let

$$n_s^\varepsilon := (dF_s)_{X_s^\varepsilon(x)} T X_s^\varepsilon L_s - F_s(X_s^\varepsilon(x)) \int_0^s \langle T X_r^\varepsilon \dot{L}_r, A(X_r^\varepsilon(x))\mathrm{d}Z_r^\varepsilon\rangle. \tag{3.13}$$

Just as in the first order case the perturbation is assumed to consist of the factorization $X_s^\varepsilon(x) = X_s \circ H_s^\varepsilon(x)$, where $H_s^\varepsilon(x)$ is given by

$$\mathrm{d}H_s^\varepsilon = X_{s*}^{-1} A(X_s \circ H_s^\varepsilon)\varepsilon k_s \mathrm{d}s,$$
$$H_s^0 = \mathrm{id}_M, \quad \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} H_0^\varepsilon(x) = v,$$

and k_s is determined by K_s via

$$k_s = A(X_s(x))^* X_{s*} \dot{K}_s.$$

Generalized Itô's formula yields $\mathrm{d}Z_s^\varepsilon = \mathrm{d}Z_s + \varepsilon k_s \mathrm{d}s$ and the appropriate Girsanov correction term again is

$$G_s^\varepsilon = \mathscr{E}\Big(-\int_0^s \varepsilon\langle k_r, \mathrm{d}Z_r\rangle\Big).$$

So we derive a second order local martingale by differentiating $\frac{\nabla}{\partial\varepsilon}\big|_{\varepsilon=0}(n_s^\varepsilon G_s^\varepsilon)$.
Mainly, the computations are the same or very similar to our first proof of Theorem 3.1 except for one difficulty: This time we need to differentiate covariantly with respect to ε at 0 the perturbed derivative process TX_s^ε which takes values in $\mathrm{Lin}(T_xM, T_{X_s^\varepsilon(x)}M)$. For this purpose we use the factorization induced by the underlying process $X_s^\varepsilon(x) = X_s \circ H_s^\varepsilon(x)$, which reads

$$TX_s^\varepsilon = TX_s \circ TH_s^\varepsilon : T_xM \to T_{H_s^\varepsilon(x)}M \to T_{X_s\circ H_s^\varepsilon(x)}M \equiv T_{X_s^\varepsilon(x)}M.$$

As we had H_s^ε as a (pathwise) solution to

$$\mathrm{d}H_s^\varepsilon = X_{s*}^{-1}A(X_s \circ H_s^\varepsilon)\varepsilon k_s \mathrm{d}s,$$

its derivative flow has to solve the formally differentiated equation

$$\mathrm{d}(TH_s^\varepsilon) = \nabla_{TH_s^\varepsilon}(X_{s*}^{-1}A(X_s \circ H_s^\varepsilon(\,\cdot\,))\varepsilon k_s)\,\mathrm{d}s.$$

We can easily verify that $\frac{\nabla}{\partial\varepsilon}\big|_{\varepsilon=0}\mathrm{d}(TH_s^\varepsilon) = \mathrm{d}\left(\frac{\nabla}{\partial\varepsilon}\big|_{\varepsilon=0}TH_s^\varepsilon\right)$ because of the linear dependence on ε in the right hand side of the upper equation, one just has to repeat the calculation of Remark 2.18 on the level of tangent spaces.
For this reason we obtain the following formula for $\frac{\nabla}{\partial\varepsilon}\big|_{\varepsilon=0}TH_s^\varepsilon \in \Gamma(T^*M\otimes TM)$:

$$\begin{aligned}\frac{\nabla}{\partial\varepsilon}\bigg|_{\varepsilon=0}TH_s^\varepsilon = \frac{\nabla}{\partial\varepsilon}\bigg|_{\varepsilon=0}TH_0^\varepsilon &- \int_0^s X_{r*}^{-1}\nabla TX_r\left(X_{r*}^{-1}A(X_r)k_r,\,\cdot\,\right)\mathrm{d}r\\ &+\int_0^s X_{r*}^{-1}\nabla A(X_{r*}(\,\cdot\,))k_r\mathrm{d}r.\end{aligned}\tag{3.14}$$

We assume $\frac{\nabla}{\partial\varepsilon}\big|_{\varepsilon=0}TH_0^\varepsilon = 0$ which only is an additional requirement for H_0^ε.
So finally we derive

$$\begin{aligned}\frac{\nabla}{\partial\varepsilon}\bigg|_{\varepsilon=0}TX_s^\varepsilon &= \frac{\nabla}{\partial\varepsilon}\bigg|_{\varepsilon=0}\left((TX_s)_{H_s^\varepsilon(x)}\right)TH_s^0 + X_{s*}\frac{\nabla}{\partial\varepsilon}\bigg|_{\varepsilon=0}TH_s^\varepsilon\\ &= \nabla TX_s\left(\frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0}H_s^\varepsilon(x),\,\cdot\,\right) - X_{s*}\int_0^s X_{r*}^{-1}\nabla TX_r\left(X_{r*}^{-1}A(X_r)k_r,\,\cdot\,\right)\mathrm{d}r\\ &\quad + X_{s*}\int_0^s X_{r*}^{-1}\nabla A(X_{r*}(\,\cdot\,))k_r\mathrm{d}r\\ &= \nabla TX_s(K_s,\,\cdot\,) - X_{s*}\int_0^s X_{r*}^{-1}\nabla TX_r\left(\dot K_r,\,\cdot\,\right)\mathrm{d}r\\ &\quad + X_{s*}\int_0^s X_{r*}^{-1}\nabla A(X_{r*}(\,\cdot\,))A(X_r(x))^*X_{r*}\dot K_r\mathrm{d}r,\end{aligned}$$

and we are able to compute

$$\begin{aligned}\frac{\nabla}{\partial\varepsilon}\bigg|_{\varepsilon=0}(n_s^\varepsilon G_s^\varepsilon) &= \frac{\nabla}{\partial\varepsilon}\bigg|_{\varepsilon=0}n_s^\varepsilon G_s^0 + n_s^0\frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0}G_s^\varepsilon\\ &= \nabla dF_s\left(X_{s*}\frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0}H_s^\varepsilon(x), X_{s*}L_s\right) + (dF_s)_{X_s(x)}\frac{\nabla}{\partial\delta}\bigg|_{\delta=0}(TX_s^\varepsilon)L_s\end{aligned}$$

$$
\begin{aligned}
&- (dF_s)_{X_s(x)} X_{s*} \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} H_s^\varepsilon(x) \int_0^s \langle X_{r*}\dot{L}_r, A(X_r(x))\mathrm{d}Z_r\rangle \\
&- F_s(X_s(x)) \int_0^s \left\langle \frac{\nabla}{\partial \varepsilon}\Big|_{\varepsilon=0} (TX_r^\varepsilon)\dot{L}_r, A(X_r(x))\mathrm{d}Z_r \right\rangle \\
&- F_s(X_s(x)) \int_0^s \left\langle X_{r*}\dot{L}_r, \nabla A(X_{r*} \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} H_r^\varepsilon(x))\mathrm{d}Z_r \right\rangle \\
&- F_s(X_s(x)) \int_0^s \langle X_{r*}\dot{L}_r, A(X_r(x))k_r\rangle \mathrm{d}r \\
&- (dF_s)_{X_s(x)} X_{s*} L_s \int_0^s \langle X_{r*}\dot{K}_r, A(X_r(x))\mathrm{d}Z_r\rangle \\
&+ F_s(X_s(x)) \int_0^s \langle X_{r*}\dot{L}_r, A(X_r(x))\mathrm{d}Z_r\rangle \int_0^s \langle X_{r*}\dot{K}_r, A(X_r(x))\mathrm{d}Z_r\rangle \\
= {}& \nabla dF_s\,(X_{s*}K_s, X_{s*}L_s) + (dF_s)_{X_s(x)} \nabla TX_s(K_s, L_s) \\
&- (dF_s)_{X_s(x)} X_{s*} \int_0^s X_{r*}^{-1} \nabla TX_r(\dot{K}_r, L_s)\mathrm{d}r \\
&+ (dF_s)_{X_s(x)} X_{s*} \int_0^s X_{r*}^{-1} \nabla A(X_{r*}L_s) A(X_r(x))^* X_{r*}\dot{K}_r \mathrm{d}r \\
&- (dF_s)_{X_s(x)} X_{s*} K_s \int_0^s \langle X_{r*}\dot{L}_r, A(X_r(x))\mathrm{d}Z_r\rangle \\
&- (dF_s)_{X_s(x)} X_{s*} L_s \int_0^s \langle X_{r*}\dot{K}_r, A(X_r(x))\mathrm{d}Z_r\rangle \\
&- F_s(X_s(x)) \int_0^s \langle \nabla TX_r(K_r, \dot{L}_r), A(X_r(x))\mathrm{d}Z_r\rangle \\
&+ F_s(X_s(x)) \int_0^s \left\langle X_{r*} \int_0^r X_{u*}^{-1} \nabla TX_u(\dot{K}_u, \dot{L}_r)\mathrm{d}u, A(X_r(x))\mathrm{d}Z_r \right\rangle \\
&- F_s(X_s(x)) \int_0^s \left\langle X_{r*} \int_0^r X_{u*}^{-1} \nabla A(X_{u*}\dot{L}_r) A(X_u(x))^* X_{u*}\dot{K}_u \mathrm{d}u, A(X_r(x))\mathrm{d}Z_r \right\rangle \\
&- F_s(X_s(x)) \int_0^s \langle X_{r*}\dot{L}_r, \nabla A(X_{r*}K_r)\mathrm{d}Z_r\rangle \\
&+ F_s(X_s(x)) \int_0^s \langle X_{r*}\dot{L}_r, A(X_r(x))\mathrm{d}Z_r\rangle \int_0^s \langle X_{r*}\dot{K}_r, A(X_r(x))\mathrm{d}Z_r\rangle \\
&- F_s(X_s(x)) \int_0^s \langle X_{r*}\dot{L}_r, X_{r*}\dot{K}_r\rangle \mathrm{d}r.
\end{aligned}
$$

It remains to carry out some integration by parts computations to derive (3.1) from this result. If we apply (2.12) with K_s replaced by $\int_0^s X_{r*}^{-1} \nabla TX_r(K_r, \dot{L}_s)\mathrm{d}r$, we get

$$
\begin{aligned}
&F_s(X_s(x)) \int_0^s \langle \nabla TX_r(K_r, \dot{L}_r), A(X_r(x))\mathrm{d}Z_r\rangle \\
&\overset{m}{=} (dF_s)_{X_s(x)} X_{s*} \int_0^s X_{r*}^{-1} \nabla TX_r(K_r, \dot{L}_r)\mathrm{d}r.
\end{aligned}
$$

Analogously we obtain

$$
F_s(X_s(x)) \int_0^s \left\langle X_{r*} \int_0^r X_{u*}^{-1} \nabla TX_u(\dot{K}_u, \dot{L}_r)\mathrm{d}u, A(X_r(x))\mathrm{d}Z_r \right\rangle
$$

$$\stackrel{m}{=} (dF_s)_{X_s(x)} X_{s*} \int_0^s \left(\int_0^r X_{u*}^{-1} \nabla T X_u(\dot{K}_u, \dot{L}_r) \mathrm{d}u \right) \mathrm{d}r$$

as well as

$$\begin{aligned} &F_s(X_s(x)) \int_0^s \left\langle X_{r*} \int_0^r X_{u*}^{-1} \nabla A(X_{u*}\dot{L}_r) A(X_u(x))^* X_{u*}\dot{K}_u \mathrm{d}u, A(X_r(x))\mathrm{d}Z_r \right\rangle \\ &\stackrel{m}{=} (dF_s)_{X_s(x)} X_{s*} \int_0^s \left(\int_0^r X_{u*}^{-1} \nabla A(X_{u*}\dot{L}_r) A(X_u(x))^* X_{u*}\dot{K}_u \mathrm{d}u \right) \mathrm{d}r \end{aligned}$$

and finally

$$\begin{aligned} &F_s(X_s(x)) \int_0^s \langle X_{r*}\dot{L}_r, \nabla A(X_{r*}K_r)(A(X_r(x)))^* A(X_r(x))\mathrm{d}Z_r \rangle \\ &\stackrel{m}{=} -(dF_s)_{X_s(x)} X_{s*} \int_0^s X_{r*}^{-1} \nabla A(X_{r*}K_r)(A(X_r(x)))^* X_{r*}\dot{L}_r \mathrm{d}r. \end{aligned}$$

So if we replace the left hand sides of these four equations by the right hand ones in the upper result for $\frac{\nabla}{\partial\varepsilon}\big|_{\varepsilon=0} (n_s^\varepsilon G_s^\varepsilon)$, we again end up with a local martingale.
After all, we use two easy integration by parts arguments of type

$$\left(\int_0^s \mathrm{d}R_r \right) (L_s) = \int_0^s \left(\int_0^r \mathrm{d}R_u \right) (\dot{L}_r) \mathrm{d}r + \int_0^s L_r \mathrm{d}R_r$$

(R an arbitrary continuous semimartingale) to derive

$$\begin{aligned} &- (dF_s)_{X_s(x)} X_{s*} \int_0^s X_{r*}^{-1} \nabla T X_r(\dot{K}_r, L_s) \mathrm{d}r \\ &+ (dF_s)_{X_s(x)} X_{s*} \int_0^s \left(\int_0^r X_{u*}^{-1} \nabla T X_u(\dot{K}_u, \dot{L}_r) \mathrm{d}u \right) \mathrm{d}r \\ &\stackrel{m}{=} -(dF_s)_{X_s(x)} X_{s*} \int_0^s X_{r*}^{-1} \nabla T X_r(\dot{K}_r, L_r) \mathrm{d}r \end{aligned}$$

and

$$\begin{aligned} &(dF_s)_{X_s(x)} X_{s*} \int_0^s X_{r*}^{-1} \nabla A(X_{r*}L_s) A(X_r(x))^* X_{r*}\dot{K}_r \mathrm{d}r \\ &- (dF_s)_{X_s(x)} X_{s*} \int_0^s \left(\int_0^r X_{u*}^{-1} \nabla A(X_{u*}\dot{L}_r) A(X_u(x))^* X_{u*}\dot{K}_u \mathrm{d}u \right) \mathrm{d}r \\ &\stackrel{m}{=} (dF_s)_{X_s(x)} X_{s*} \int_0^s X_{r*}^{-1} \nabla A(X_{r*}L_r) A(X_r(x))^* X_{r*}\dot{K}_r \mathrm{d}r \end{aligned}$$

and including these equalities (modulo local martingales) into our computations up to now, we obtain the claimed equation (3.1). □

3.1.2. Main result and shorter proofs

Of course one could now formulate a theorem that takes expectations of (3.1) at times 0 and t under conditions such that it is a true martingale. In the semigroup case for example, this yields a stochastic representation of $\nabla dP_t f(v, w)$ with terms that are perfectly symmetric in K_s and L_s, the first one starting in v and the second one in w.
For applications this result is too complicated, and we find a shorter formula if we additionally assume that K_s and L_s vary on disjoint (stochastic) time intervals.

Theorem 3.4 (Non-intrinsic representation formula for the Hessian).
Fix $x \in M$ and let $0 < \widetilde{\sigma} < \widetilde{\tau}$ denote the first exit times of $X(x)$ from two open relatively compact neighbourhoods $D_1 \subset \bar{D}_1 \subset D_2$ of x. We fix $t > 0$ and write $\sigma := \widetilde{\sigma} \wedge \frac{t}{2}$ as well as $\tau := \widetilde{\tau} \wedge t$.
In addition to this, let $f : M \to \mathbb{R}$ be bounded measurable and $v, w \in T_xM$. The bounded adapted processes $(K_s)_{s\in[0,t]}$ and $(L_s)_{s\in[0,t]}$ with paths in $\mathbb{H}([0,t], T_xM)$ are assumed to fulfill the following properties:

i) $K_0 = v$, $K_t = 0$, $\dot{K}_s = 1_{]\sigma,\tau]}(s)\dot{K}_s$ *(hence $K_s = v$ for $s \le \sigma$ and then decays to 0 within $]\sigma, \tau]$).*

ii) $L_0 = w$, $L_t = 0$, $\dot{L}_s = 1_{[0,\sigma]}(s)\dot{L}_s$ *(and therefore $L|_{[\sigma,\tau]} = 0$).*

iii) $\int_0^t \|\dot{K}_s\|^2 \mathrm{d}s, \int_0^t \|\dot{L}_s\|^2 \mathrm{d}s \in L^{1+\alpha}(\mathbb{P})$ *for some $\alpha > 0$ (observe the difference in the order of integrability compared to Theorem 2.19).*

Under these assumptions we derive the following stochastic representation formula for the Hessian $\mathrm{Hess}_x F_0(v, w) \equiv \nabla dF_0(v, w)$*:*

$$\begin{aligned}\nabla dF_0(v,w) = &-\mathbb{E}\left[F_\sigma(X_\sigma(x)) \int_0^\sigma \langle \nabla TX_r(v, \dot{L}_r), A(X_r(x))\mathrm{d}Z_r\rangle\right] \\ &- \mathbb{E}\left[F_\sigma(X_\sigma(x)) \int_0^\sigma \langle X_{r*}\dot{L}_r, \nabla A(X_{r*}v)\mathrm{d}Z_r\rangle\right] \\ &+ \mathbb{E}\left[F_\tau(X_\tau(x)) \int_0^\sigma \langle X_{r*}\dot{L}_r, A(X_r(x))\mathrm{d}Z_r\rangle\right. \\ &\qquad \left.\int_\sigma^\tau \langle X_{r*}\dot{K}_r, A(X_r(x))\mathrm{d}Z_r\rangle\right].\end{aligned} \tag{3.15}$$

In the heat semigroup case the result modifies to

$$\begin{aligned}\nabla dP_tf(v,w) = &-\mathbb{E}\left[f(X_t(x))1_{\{t<\zeta(x)\}} \int_0^\sigma \langle \nabla TX_r(v, \dot{L}_r), A(X_r(x))\mathrm{d}Z_r\rangle\right] \\ &- \mathbb{E}\left[f(X_t(x))1_{\{t<\zeta(x)\}} \int_0^\sigma \langle X_{r*}\dot{L}_r, \nabla A(X_{r*}v)\mathrm{d}Z_r\rangle\right] \\ &+ \mathbb{E}\left[f(X_t(x))1_{\{t<\zeta(x)\}} \int_0^\sigma \langle X_{r*}\dot{L}_r, A(X_r(x))\mathrm{d}Z_r\rangle\right. \\ &\qquad \left.\int_\sigma^\tau \langle X_{r*}\dot{K}_r, A(X_r(x))\mathrm{d}Z_r\rangle\right].\end{aligned} \tag{3.16}$$

We discuss briefly how the results of [E-L 1] on the Hessian of heat semigroups are contained in our formulae.

Remark 3.5. In the case of compact M (such that $\zeta = \infty$) and $f \in bC^2(M)$ for $\sigma = \frac{t}{2}$ as well as $\tau = t$, we could take $K_s = v - \frac{2}{t}((s - \frac{t}{2}) \vee 0)v$ and $L_s = w - \frac{2}{t}(s \wedge \frac{t}{2})w$ to derive

$$\nabla dP_tf(v,w) = \frac{4}{t^2}\mathbb{E}\left[f(X_t(x)) \int_0^{t/2} \langle X_{r*}w, A(X_r(x))\mathrm{d}Z_r\rangle \int_{t/2}^t \langle X_{r*}v, A(X_r(x))\mathrm{d}Z_r\rangle\right]$$

$$+\frac{2}{t}\mathbb{E}\left[f(X_t(x))\int_0^{t/2}\langle\nabla TX_r(v,w),A(X_r(x))\mathrm{d}Z_r\rangle\right]$$
$$+\frac{2}{t}\mathbb{E}\left[f(X_t(x))\int_0^{t/2}\langle X_{r*}w,\nabla A(X_{r*}v)\mathrm{d}Z_r\rangle\right].$$

This is almost the same as Elworthy-Li's result in [E-L 1], Theorem 3.1, except for their permutation of w and v in the first term of the right hand side. This is due to a slight error in the proof of Theorem 2.3 in that paper. One could also use the attempt of Remark 3.3 to use processes K_s and L_s both starting at 0 and varying to v and w within $[0,\frac{t}{2}]$ and $]\frac{t}{2},t]$ respectively to derive

$$\nabla dP_tf(v,w)=\frac{4}{t^2}\mathbb{E}\left[f(X_t(x))\int_0^{t/2}\langle X_{r*}v,A(X_r(x))\mathrm{d}Z_r\rangle\int_{t/2}^{t}\langle X_{r*}w,A(X_r(x))\mathrm{d}Z_r\rangle\right]$$
$$+\frac{2}{t}\mathbb{E}\left[f(X_t(x))\int_{t/2}^{t}\langle\nabla TX_r(v,w),A(X_r(x))\mathrm{d}Z_r\rangle\right]$$
$$+\frac{2}{t}\mathbb{E}\left[f(X_t(x))\int_{t/2}^{t}\langle X_{r*}w,\nabla A(X_{r*}v)\mathrm{d}Z_r\rangle\right]$$

which now differs from the Elworthy-Li result by integration over the second time period in the last two terms instead of integrating over the first part.

Proof of Theorem 3.4. We apply our prerequisites on K_s and L_s to reduce the local martingale n_s from (3.1) to the form

$$\begin{aligned}n_s&=\nabla dF_s(X_{s*}K_s,X_{s*}L_s)+(dF_s)_{X_s(x)}\nabla TX_s(K_s,L_s)\\&-(dF_s)_{X_s(x)}X_{s*}\int_0^{\sigma\wedge s}X_{r*}^{-1}\nabla TX_r(v,\dot L_r)\mathrm{d}r\\&+(dF_s)_{X_s(x)}X_{s*}\int_0^{\sigma\wedge s}X_{r*}^{-1}\nabla A(X_{r*}v)A(X_r(x))^*X_{r*}\dot L_r\mathrm{d}r\\&-(dF_s)_{X_s(x)}X_{s*}K_s\int_0^{\sigma\wedge s}\langle X_{r*}\dot L_r,A(X_r(x))\mathrm{d}Z_r\rangle\\&+F_s(X_s(x))\int_{\sigma\wedge s}^{\tau\wedge s}\langle X_{r*}\dot K_r,A(X_r(x))\mathrm{d}Z_r\rangle\int_0^{\sigma\wedge s}\langle X_{r*}\dot L_r,A(X_r(x))\mathrm{d}Z_r\rangle.\end{aligned}$$

because integrals over $\dot K$ need only be taken from time σ on, and this implies that L already has vanished.

Application of Theorem 2.14 for the pathwise integrals $\int_0^s X_{r*}^{-1}\nabla TX_r(v,\dot L_r)\mathrm{d}r$ as well as $\int_0^s X_{r*}^{-1}\nabla A(X_{r*}v)A(X_r(x))^*X_{r*}\dot L_r\mathrm{d}r$ instead of K_s leads to the local martingale N_s (because it differs from n_s only by two local martingales of first order type)

$$\begin{aligned}N_s&=\nabla dF_s(X_{s*}K_s,X_{s*}L_s)+(dF_s)_{X_s(x)}\nabla TX_s(K_s,L_s)\\&-F_s(X_s(x))\int_0^{\sigma\wedge s}\langle\nabla TX_r(v,\dot L_r),A(X_r(x))\mathrm{d}Z_r\rangle\end{aligned}$$

$$
\begin{aligned}
&- F_s(X_s(x)) \int_0^{\sigma\wedge s} \langle X_{r*}\dot{L}_r, \nabla A(X_{r*}v)\mathrm{d}Z_r\rangle \\
&- (dF_s)_{X_s(x)} X_{s*}K_s \int_0^{\sigma\wedge s} \langle X_{r*}\dot{L}_r, A(X_r(x))\mathrm{d}Z_r\rangle \\
&+ F_s(X_s(x)) \int_{\sigma\wedge s}^{\tau\wedge s} \langle X_{r*}\dot{K}_r, A(X_r(x))\mathrm{d}Z_r\rangle \int_0^{\sigma\wedge s} \langle X_{r*}\dot{L}_r, A(X_r(x))\mathrm{d}Z_r\rangle
\end{aligned}
$$

(we again made use of $\nabla A(X_{r*}v)^* A(X_r(x)) = -A(X_r(x))^*\nabla A(X_{r*}v)$).
In the following, we verify that this new local martingale N_s is a true one on $[0,\tau]$ by verifying that $\{N_\varrho : 0 \le \varrho \le \tau \text{ stopping time}\}$ is uniformly integrable.
Due to the continuity of F and its derivatives on the compact set $[0,t]\times\bar{D}_2$, we are given finite constants $C_0 := \sup_{(s,y)\in[0,t]\times\bar{D}_2} |F_s(y)|$ as well as $C_1 := \sup_{(s,y)\in[0,t]\times\bar{D}_2} \|d(F_s)_y\|$ and finally $C_2 := \sup_{(s,y)\in[0,t]\times\bar{D}_2} \|\nabla d(F_s)_y\|$ (maximum norm of a bilinear map on $T_xM\times T_xM$).
As in the proof of Theorem 2.19 we make use of $g_1 := \sup_{s\in[0,\tau]} \|T_xX_s\| \in L^p(\mathbb{P})$ as well as $g_2 := \sup_{s\in[0,\tau]} \|\nabla T_xX_s\| \in L^p(\mathbb{P})$ for arbitrary $1\le p<\infty$ (because this holds for diffusions on compact manifolds and we may modify M outside D_2).
By choosing upper bounds C_4 for $|K_s|$ and C_5 for $|L_s|$ on $[0,\tau]$ we get for any $0\le\varrho\le\tau$

$$
\begin{aligned}
|N_\varrho^\tau| \le & C_2 g_1^2 C_4 C_5 + C_1 g_2 C_4 C_5 \\
&+ C_0 \left| \int_0^{\sigma\wedge\varrho} \langle A(X_r(x))^* \nabla T X_r(v, \dot{L}_r), \mathrm{d}Z_r\rangle \right| \\
&+ C_0 \left| \int_0^{\sigma\wedge\varrho} \langle \nabla A(X_{r*}v)^* X_{r*}\dot{L}_r, \mathrm{d}Z_r\rangle \right| \\
&+ C_1 g_1 C_4 \left| \int_0^{\sigma\wedge\varrho} \langle A(X_r(x))^* X_{r*}\dot{L}_r, \mathrm{d}Z_r\rangle \right| \\
&+ C_0 \left| \int_{\sigma\wedge\varrho}^{\tau\wedge\varrho} \langle A(X_r(x))^* X_{r*}\dot{K}_r, \mathrm{d}Z_r\rangle \right| \left| \int_0^{\sigma\wedge\varrho} \langle A(X_r(x))^* X_{r*}\dot{L}_r, \mathrm{d}Z_r\rangle \right| .
\end{aligned}
\tag{3.17}
$$

The first two terms of the right hand side have finite expectation since g_1 and g_2 raised to any power are integrable. Moreover, the same Burkholder-Gundy and Hölder argument as in the proof of 2.19 provides that for some $c>0$

$$
\begin{aligned}
\mathbb{E}\left| \int_0^\sigma \langle A(X_r(x))^*\nabla T X_r(v,\dot{L}_r), \mathrm{d}Z_r\rangle \right| &\le c\mathbb{E}\left(\int_0^\sigma \|\nabla T X_r(v,\dot{L}_r)\|^2 \,\mathrm{d}r \right)^{\frac{1}{2}} \\
&\le c\left[\mathbb{E}\left(\sup_{0\le r\le\sigma} \|\nabla T X_r(v,\,\cdot\,)\| \right)^{\frac{1+\alpha}{\alpha}} \right]^{\frac{\alpha}{1+\alpha}} \left[\mathbb{E}\left(\int_0^\sigma \|\dot{L}_r\|^2\,\mathrm{d}r \right)^{\frac{1+\alpha}{2}} \right]^{\frac{1}{1+\alpha}} < \infty,
\end{aligned}
$$

because $\sup_{0\le r\le\tau} \|\nabla T X_r(v,\,\cdot\,)\| \le g_2\|v\|$. Herein we only made use of half the order of integrability of $\int_0^\tau \|\dot{L}_r\|^2\,\mathrm{d}r$ assumed in the theorem.
The analogous computation yields

$$
\mathbb{E}\left| \int_0^\sigma \langle \nabla A(X_{r*}v)^* X_{r*}\dot{L}_r, \mathrm{d}Z_r\rangle \right| \le c\mathbb{E}\left(\int_0^\sigma \|\nabla A(X_{r*}v)^* X_{r*}\dot{L}_r\|^2\,\mathrm{d}r \right)^{\frac{1}{2}}
$$

$$\leq c\left[\mathbb{E}\left(\sup_{0\leq r\leq\sigma}\|\nabla A(X_{r*}v)^*X_{r*}(\,\cdot\,)\|\right)^{\frac{1+\alpha}{\alpha}}\right]^{\frac{\alpha}{1+\alpha}}\left[\mathbb{E}\left(\int_0^\sigma\|\dot{L}_r\|^2\,\mathrm{d}r\right)^{\frac{1+\alpha}{2}}\right]^{\frac{1}{1+\alpha}}.$$

The right hand side is finite, if we have $\sup_{0\leq r\leq\tau}\|\nabla A(X_{r*}v)^*X_{r*}(\,\cdot\,)\|\in L^{(1+\alpha)/\alpha}(\mathbb{P})$. But as $\nabla A(\,\cdot\,)^*(\,\cdot\,)\in \mathrm{Bil}(TM,TM;\mathbb{R}^m)$ is bounded on the compact set D_2 we find that this sup is less or equal $C_6\|v\|g_2^2$ (with C_6 just depending on A) and the integrability of g_2 to any power is sufficient for this part.
For the fifth term of the right hand side of (3.17) according to the Schwartz inequality

$$\mathbb{E}\left(g_1\left|\int_0^{\sigma\wedge\varrho}\langle A(X_r(x))^*X_{r*}\dot{L}_r,\mathrm{d}Z_r\rangle\right|\right)\leq\|g_1\|_2\left(\mathbb{E}\left|\int_0^{\sigma\wedge\varrho}\langle A(X_r(x))^*X_{r*}\dot{L}_r,\mathrm{d}Z_r\rangle\right|^2\right)^{1/2}.$$

Again by Burkholder-Gundy we compute

$$\mathbb{E}\left|\int_0^{\sigma\wedge\varrho}\langle A(X_r(x))^*X_{r*}\dot{L}_r,\mathrm{d}Z_r\rangle\right|^2\leq c\mathbb{E}\int_0^\sigma\|X_{r*}\dot{L}_r\|^2\,\mathrm{d}r\leq c\mathbb{E}\left(g_1^2\int_0^\sigma\|\dot{L}_r\|^2\,\mathrm{d}r\right).$$

Obviously, by another Hölder argument with respect to the partition $1=\frac{\alpha}{1+\alpha}+\frac{1}{1+\alpha}$ we split the right hand side into two expectations and the assumption $\int_0^t\|\dot{L}_r\|^2\,\mathrm{d}r$ in $L^{1+\alpha}(\mathbb{P})$ provides that this is finite.
Now we focus on the last term for which Hölder yields

$$\begin{aligned}&\mathbb{E}\left[\left|\int_{\sigma\wedge\varrho}^{\tau\wedge\varrho}\langle A(X_r(x))^*X_{r*}\dot{K}_r,\mathrm{d}Z_r\rangle\right|\left|\int_0^{\sigma\wedge\varrho}\langle A(X_r(x))^*X_{r*}\dot{L}_r,\mathrm{d}Z_r\rangle\right|\right]\\&\leq\left(\mathbb{E}\left|\int_{\sigma\wedge\varrho}^{\tau\wedge\varrho}\langle A(X_r(x))^*X_{r*}\dot{K}_r,\mathrm{d}Z_r\rangle\right|^2\right)^{1/2}\left(\mathbb{E}\left|\int_0^{\sigma\wedge\varrho}\langle A(X_r(x))^*X_{r*}\dot{L}_r,\mathrm{d}Z_r\rangle\right|^2\right)^{1/2}\end{aligned}$$

and reasoning as in the case just before proves that both factors are finite. So $(N_s)_{0\leq s\leq\tau}$ is a true martingale.
Comparing expectations for $s=0$ and $s=\tau$ yields the Bismut formula (3.15), where we used $\nabla TX_0=0$ since $X_0=\mathrm{id}_M$ and the martingale property of $F_s\circ X_s(x)$ for $s\leq\tau$ to derive $\mathbb{E}[F_\tau\circ X_\tau(x)|\mathscr{F}_\sigma^Z]=\mathbb{E}[F_\sigma\circ X_\sigma(x)]$.
In particular, for the Hessian of the heat semigroup P_tf we have the preliminary result for $\tau=\tilde{\tau}\wedge(t-\varepsilon)$ and $K_s\equiv 0$ on $[t-\varepsilon,t]$, where $\varepsilon<\frac{t}{2}$:

$$\begin{aligned}\nabla d(P_tf)(v,w)=&-\mathbb{E}\left[P_{t-\tau}f(X_\tau(x))\int_0^\sigma\langle\nabla TX_r(v,\dot{L}_r),A(X_r(x))\mathrm{d}Z_r\rangle\right]\\&-\mathbb{E}\left[P_{t-\tau}f(X_\tau(x))\int_0^\sigma\langle X_{r*}\dot{L}_r,\nabla A(X_{r*}v)\mathrm{d}Z_r\rangle\right]\\&+\mathbb{E}\left[P_{t-\tau}f(X_\tau(x))\int_\sigma^\tau\langle X_{r*}\dot{K}_r,A(X_r(x))\mathrm{d}Z_r\rangle\right.\\&\qquad\left.\int_0^\sigma\langle X_{r*}\dot{L}_r,A(X_r(x))\mathrm{d}Z_r\rangle\right].\end{aligned}$$

Application of the strong Markov property $\mathbb{E}[f\circ X_t(x)1_{\{t<\zeta(x)\}}|\mathscr{F}_\tau^Z]=P_{t-\tau}f(X_\tau(x))$ yields (3.16) for K_s vanishing from $\tilde{\tau}\wedge(t-\varepsilon)$ up to t.
The extension to general finite energy processes K_s works as in the first order case by approximating with suitable K_s^ε. □

If our goal for second order derivatives was just to prove Theorem 3.4, we could replace the computations in the preceding section by the shorter following argument:

Theorem 3.6. *Fix $x \in M$ and $v, w \in T_xM$. Let (L_s) be a $(\mathscr{F}_s^Z)$-adapted process with sample paths in $\mathbb{H}([0,t], T_xM)$ and $L_0 = w$. Then we have the local martingale*

$$\begin{aligned} n_s := {}& \nabla dF_s(X_{s*}v, X_{s*}L_s) + (dF_s)_{X_s(x)} \nabla TX_s(v, L_s) \\ &- (dF_s)_{X_s(x)} X_{s*} \int_0^s X_{r*}^{-1} \nabla TX_r(v, \dot{L}_r) \mathrm{d}r \\ &+ (dF_s)_{X_s(x)} X_{s*} \int_0^s X_{r*}^{-1} \nabla A(X_{r*}v) A(X_r(x))^* X_{r*} \dot{L}_r \mathrm{d}r \\ &- (dF_s)_{X_s(x)} X_{s*} v \int_0^s \langle X_{r*} \dot{L}_r, A(X_r(x)) \mathrm{d}Z_r \rangle. \end{aligned} \tag{3.18}$$

Obviously, this result is formula (3.1) in the particular case of $k_s = 0$ and therefore all terms depending linearly on $\dot{K}_s$ vanish. In other words, one drops the dependence on ε the except for the deterministic one of the initial value $H_0^{\varepsilon,0}(x)$ fulfilling $\frac{\partial}{\partial \varepsilon}\big|_{\varepsilon=0} H_0^{\varepsilon,0}(x) = v = K_0 \equiv K_s$, $0 \le s \le t$. We give a short proof based on the observation that in this special case we can simply take the covariant derivative in direction v of the first order local martingale (depending on L_s instead of K_s).

Proof of Theorem 3.6. We compute the covariant differential in direction v of the first order local martingale from (2.12) (with K_s replaced by L_s)

$$\begin{aligned} &\nabla_v \left[(dF_s)_{X_s(x)} X_{s*} L_s - F_s(X_s(x)) \int_0^s \langle X_{r*} \dot{L}_r, A(X_r(x)) \mathrm{d}Z_r \rangle \right] \\ &= \nabla dF_s(X_{s*}v, X_{s*}L_s) + (dF_s)_{X_s(x)} \nabla TX_s(v, L_s) \\ &\quad - (dF_s)_{X_s(x)} X_{s*} v \int_0^s \langle X_{r*} \dot{L}_r, A(X_r(x)) \mathrm{d}Z_r \rangle \\ &\quad - F_s(X_s(x)) \int_0^s \langle \nabla TX_r(v, \dot{L}_r), A(X_r(x)) \mathrm{d}Z_r \rangle \\ &\quad - F_s(X_s(x)) \int_0^s \langle X_{r*} \dot{L}_r, \nabla A(X_{r*}v) \mathrm{d}Z_r \rangle, \end{aligned}$$

which is again a local martingale. The step to the claimed formula consists of the same integration by parts arguments as in the proof of 3.4 above (where we derived n_s from N_s). □

Remark 3.7. It should be obvious that and which way it is possible to prove Theorem 3.4 by means of this local martingale instead of (3.1) except for the last term.
For that purpose under the assumptions of 3.4 we consider the two local martingales

$$\begin{aligned} O_s^1 &:= (dF_s)_{X_s(x)} X_{s*}(K_s - v) - F_s \circ X_s(x) \int_0^s \langle X_{r*} \dot{K}_r, A(X_r(x)) \mathrm{d}Z_r \rangle, \\ O_s^2 &:= \int_0^s \langle X_{r*} \dot{L}_r, A(X_r(x)) \mathrm{d}Z_r \rangle. \end{aligned}$$

We obtain $\mathrm{d}(O^1_s O^2_s) \overset{m}{=} \mathrm{d}O^1_s \mathrm{d}O^2_s = 0$ because of $O^1 = 0$ on $[0,\sigma]$ and O^2 being constant on $]\sigma,\tau]$. According to the assumptions on K and L, (again by a Hölder and Burkholder-Gundy argument) O^1O^2 is a true martingale on $[0,\tau]$, which provides $\mathbb{E}[M_\tau O_\tau] = 0$ and therefore, as $K_\tau = 0$,

$$\begin{aligned}\mathbb{E}&\left[(dF_\tau)_{X_\tau(x)} X_{\tau *} v \int_0^\sigma \langle X_{r*}\dot{L}_r, A(X_r(x))\mathrm{d}Z_r\rangle\right] \\ &= -\mathbb{E}\left[F_\tau \circ X_\tau(x) \int_\sigma^\tau \langle X_{r*}\dot{K}_r, A(X_r(x))\mathrm{d}Z_r\rangle \int_0^\sigma \langle X_{r*}\dot{L}_r, A(X_r(x))\mathrm{d}Z_r\rangle\right].\end{aligned} \tag{3.19}$$

3.1.3. Integration by parts argument

In contrast to the first order case where the integration by parts method provides the shortest proof of the derivative formula, in the second order case these proofs do not seem to be suitable to derive the most general local martingale (3.1).
However, we are able to give a purely integration by parts based proof of Theorem 3.6 right before.

Second proof of Theorem 3.6. Consider the local martingale m_s taking values in T_xM^* given by

$$m_s w := \nabla dF_s(X_{s*}v, X_{s*}w) + (dF_s)_{X_s(x)} \nabla TX_s(v,w)$$

for $v, w \in T_xM$ (cf. Remark 3.3). If we replace w in here by the process L_s from above, Itô's integration by parts formula tells us that $m_s L_s = \int_0^s m_r \dot{L}_r \mathrm{d}r + \int_0^s (\mathrm{d}m_r)L_r$, so we have the local martingale

$$\begin{aligned}\widetilde{m}_s &:= \nabla dF_s(X_{s*}v, X_{s*}L_s) + (dF_s)_{X_s(x)} \nabla TX_s(v, L_s) \\ &\quad - \int_0^s \nabla dF_r(X_{r*}v, X_{r*}\dot{L}_r)\mathrm{d}r - \int_0^s (dF_r)_{X_r(x)} \nabla TX_r(v, \dot{L}_r)\mathrm{d}r.\end{aligned} \tag{3.20}$$

For the last term on the right hand side we put $H_s := \int_0^s X_{r*}^{-1} \nabla TX_r(v, \dot{L}_r)\mathrm{d}r$ and use again integration by parts to compute

$$\int_0^s (dF_r)_{X_r(x)} X_{r*} \dot{H}_r \mathrm{d}r \overset{m}{=} (dF_s)_{X_s(x)} X_{s*} H_s. \tag{3.21}$$

So we need a possibility to pass over from the integral over ∇dF_r to some expression where the Hessian stands outside the integral. This is achieved by applying the product rule to the local martingale $(dF_s)_{X_s(x)} X_{s*}$ with values in T_xM^*. We have

$$\begin{aligned}\mathrm{d}\big((dF_s)_{X_s(x)} X_{s*}\big) &= \mathrm{d}\left[\big((dF_s)_{X_s(x)} //_{0,s}\big)\big(//_{0,s}^{-1} X_{s*}\big)\right] \\ &= \nabla dF_s(\mathrm{d}X_s, //_{0,s}\,\cdot\,) //_{0,s}^{-1} X_{s*} + (dF_s)_{X_s(x)} //_{0,s} \mathrm{d}\big(//_{0,s}^{-1} X_{s*}\big) + \text{drift} \\ &= \nabla dF_s(\mathrm{d}X_s, X_{s*}\,\cdot\,) + (dF_s)_{X_s(x)} DTX_s + \text{drift},\end{aligned}$$

and consequently, according to the martingale parts of the stochastic differential equations (1.4) for $\mathrm{d}X_s$ and (2.3) for DTX_s

$$\mathrm{d}\big((dF_s)_{X_s(x)} X_{s*}\big) = \nabla dF_s(A(X_s(x))\mathrm{d}Z_s, X_{s*}\,\cdot\,) + (dF_s)_{X_s(x)} \nabla A(X_{s*}\,\cdot\,)\mathrm{d}Z_s \tag{3.22}$$

For this reason and writing again $\ell_s = A(X_s(x))^* X_{s*}\dot{L}_s$ we find

$$\begin{aligned}\nabla dF_s(X_{s*}v, X_{s*}\dot{L}_s)\mathrm{d}s &\overset{m}{=} \nabla dF_s(A(X_s(x))\mathrm{d}Z_s, X_{s*}v)\langle \ell_s, \mathrm{d}Z_s\rangle \\ &= \mathrm{d}\left((dF_s)_{X_s(x)}X_{s*}v\right)\langle \ell_s, \mathrm{d}Z_s\rangle - (dF_s)_{X_s(x)}\nabla A(X_{s*}v)\mathrm{d}Z_s\langle \ell_s, \mathrm{d}Z_s\rangle \\ &\overset{m}{=} \mathrm{d}\left((dF_s)_{X_s(x)}X_{s*}v\right)\langle X_{s*}\dot{L}_s, A(X_s(x))\mathrm{d}Z_s\rangle \\ &\quad - (dF_s)_{X_s(x)}\nabla A(X_{s*}v)A(X_s(x))^* X_{s*}\dot{L}_s\mathrm{d}s.\end{aligned}$$

Integration over $[0, s]$ leads to

$$\begin{aligned}\int_0^s \nabla dF_r(X_{r*}v, X_{r*}\dot{L}_r)\mathrm{d}r &\overset{m}{=} (dF_s)_{X_s(x)}X_{s*}v \int_0^s \langle X_{r*}\dot{L}_r, A(X_r(x))\mathrm{d}Z_r\rangle \\ &\quad - \int_0^s (dF_r)_{X_r(x)}\nabla A(X_{r*}v)A(X_r(x))^* X_{r*}\dot{L}_r\mathrm{d}r.\end{aligned} \tag{3.23}$$

So after another application of (3.21) now for $H_s = \int_0^s X_{r*}^{-1}\nabla A(X_{r*}v)A(X_r(x))^* X_{r*}\dot{L}_r\mathrm{d}r$ it is shown that $\widetilde{m}_s$ differs from (3.18) again only by some local martingale. □

As mentioned above, we actually did not find a proof of Theorem 3.1 completely based on integration by parts methods to obtain the most general local martingale (3.1). We will briefly describe the point of failure.

Remark 3.8. Obviously the attempt is to replace v in Theorem 3.6 by K_s the same way as done with w and L_s in the proof right before. For this purpose, we read n_s in equation (3.18) as $n_s(v)$ with the local martingale n_s taking values in T_x^*M. By $n_s(K_s) = \int_0^s (\mathrm{d}n_r)(K_r) + \int_0^s n_r(\dot{K}_r)\mathrm{d}r$ we conclude that $n_s(K_s) - \int_0^s n_r(\dot{K}_r)\mathrm{d}r$ is again a local martingale. But the problem about this is that one ends up with a term

$$-\int_0^s \nabla dF_r(X_{r*}\dot{K}_r, X_{r*}L_r)\mathrm{d}r$$

and there seems to be no way to get rid of this type of integral over ∇dF_r by generalizing the method that provided (3.23).

In some way this observation is reasonable because that procedure could be used to obtain an intrinsic general second order local martingale below (general here means depending bilinearly on two finite energy processes K_s and L_s which do not necessarily vary on disjoint time intervals only) if one starts out by intrinsic terms and filters out noise afterwards. But actually, the two Girsanov proofs above do not provide such a general intrinsic local martingale and it seems quite uncertain whether it should exist at all (at least in a comparable form to (3.1)). We will return to this point in the next paragraph.

3.2. Intrinsic local martingales for second order derivatives

As in the treatment of intrinsic first order derivatives we focus on the case that ∇ satisfies the Le Jan-Watanabe property from Remark 2.3.
Moreover, instead of working with a diffusion process with drift, we only consider the special case of a M-valued Brownian motion, i.e. $L = \frac{1}{2}\Delta$.

In particular we have (cf. Paragraph 2.1.2)

$$\operatorname{trace}(\nabla A \otimes A) = 0$$

and this combined with (2.3) defining the derivative process, which in case of Brownian motion ($V = 0$) reads

$$DTX_s = \nabla_{TX_s} A(X_s)\,\mathrm{d}Z_s - \frac{1}{2}\operatorname{Ric}^\sharp(TX_s)\,\mathrm{d}s,$$

has the immediate consequence that all terms depending bilinearly on DTX and $\mathrm{d}X$ or $\nabla_{TX}\mathrm{d}X$ and $\mathrm{d}X$ are vanishing, so

$$DTX_s \otimes \mathrm{d}X_s = \operatorname{trace}(\nabla_{TX_s} A \otimes A)(X_s)\mathrm{d}s = 0 \tag{3.24}$$

and

$$\nabla_{TX_s}\mathrm{d}X_s \otimes \mathrm{d}X_s = \operatorname{trace}(\nabla_{TX_s} A \otimes A)(X_s)\mathrm{d}s = 0. \tag{3.25}$$

The easiest possibility to derive an intrinsic second order local martingale is to differentiate the intrinsic first order martingale of equation (2.34) covariantly in direction v and filter out redundant noise. This way we obtain:

Lemma 3.9. *Keeping the notations of the foregoing paragraph and W_s and denoting the damped (or Dohrn-Guerra) transport along the paths of $X_s(x)$, the following process is a $\big(\mathscr{F}_t^{X(x)}\big)$-adapted real valued local martingale.*

$$\begin{aligned} n_s := {} & \nabla dF_s(W_s v, W_s L_s) + (dF_s)_{X_s(x)} \widetilde{\nabla W}_s(v, L_s) \\ & - (dF_s)_{X_s(x)} W_s \int_0^s W_r^{-1} \widetilde{\nabla W}_r(v, \dot{L}_r)\mathrm{d}r \\ & - (dF_s)_{X_s(x)} W_s v \int_0^s \langle W_r \dot{L}_r, A(X_r(x))\mathrm{d}Z_r\rangle, \end{aligned} \tag{3.26}$$

where $\widetilde{\nabla W}_s := \mathbb{E}[\nabla W_s | \mathscr{F}_s^{X(x)}] := //_{0,s}\mathbb{E}[//_{0,s}^{-1}\nabla W_s | \mathscr{F}_s^{X(x)}]$ is the conditional expectation of the covariant derivative of the damped transport with respect to the smaller filtration $(\mathscr{F}_s^{X(x)})$.

Proof. We repeat the proof of Theorem 3.6, but start out with the intrinsic local martingale from (2.34). So we derive

$$\begin{aligned} & \nabla_v \left[(dF_s)_{X_s(x)} W_s L_s - F_s(X_s(x)) \int_0^s \langle W_r \dot{L}_r, A(X_r(x))\mathrm{d}Z_r\rangle \right] \\ & = \nabla dF_s(X_{s*}v, W_s L_s) + (dF_s)_{X_s(x)} \nabla W_s(v, L_s) \\ & \quad - (dF_s)_{X_s(x)} X_{s*}v \int_0^s \langle W_r \dot{L}_r, A(X_r(x))\mathrm{d}Z_r\rangle \\ & \quad - F_s(X_s(x)) \int_0^s \langle \nabla W_r(v, \dot{L}_r), A(X_r(x))\mathrm{d}Z_r\rangle \\ & \quad - F_s(X_s(x)) \int_0^s \langle W_r \dot{L}_r, \nabla A(X_{r*}v)\mathrm{d}Z_r\rangle, \end{aligned}$$

which is again a local martingale. Now observe that every term of the right hand side depends linearly on just one factor that is not adapted to $\big(\mathscr{F}_s^{X(x)}\big)$. So taking conditional expectation with respect to this filtration yields the local $\big(\mathscr{F}_s^{X(x)}\big)$-martingale

$$\begin{aligned}&\nabla dF_s(W_s v, W_s L_s) + (dF_s)_{X_s(x)}\widetilde{\nabla W_s}(v, L_s)\\&- (dF_s)_{X_s(x)}W_s v\int_0^s\langle W_r\dot L_r, A(X_r(x))\mathrm{d}Z_r\rangle\\&- F_s(X_s(x))\int_0^s\langle\widetilde{\nabla W_r}(v,\dot L_r), A(X_r(x))\mathrm{d}Z_r\rangle\\&- F_s(X_s(x))\int_0^s\langle W_r\dot L_r, \nabla A(W_r v)\mathrm{d}Z_r\rangle.\end{aligned}$$

Two integration by parts arguments based on the local martingale property of (2.34) allow us to replace the last two terms of the upper result (modulo local martingales) by

$$\begin{aligned}&-(dF_s)_{X_s(x)}W_s\int_0^s W_r^{-1}\widetilde{\nabla W_r}(v,\dot L_r)\mathrm{d}r\\&+(dF_s)_{X_s(x)}W_s\int_0^s W_r^{-1}\nabla A(W_r v)A(X_r(x))^*W_r\dot L_r\mathrm{d}r.\end{aligned}$$

It remains to show that the last term depending on $\nabla A(W_r v)A(X_r(x))^*W_r\dot L_r$ vanishes. For this purpose, we consider the orthogonal decomposition of the trivial bundle $M\times\mathbb{R}^m = M\times(\ker A(\,\cdot\,)\oplus(\ker A(\,\cdot\,))^\perp)$ along the paths of $X_{\boldsymbol{.}}(x)$.
With respect to this we have $A(X_s(x))^*W_s\dot L_s = \ell_s = \ell_s^0 + \ell_s^\perp$ with $\ell_s^0\in\ker A(X_s(x))$ and $\ell_s^\perp\in(\ker A(X_s(x)))^\perp$. But then $A(X_s(x))\ell_s^\perp = A(X_s(x))\ell_s = W_s\dot L_s$ and therefore $A(X_s(x))^*W_s\dot L_s = A(X_s(x))^*A(X_s(x))\ell_s^\perp = \ell_s^\perp$ since $A(X_s(x))^*A(X_s(x))$ is the orthogonal projection on $(\ker A(X_s(x)))^\perp$.
On the other hand, for the Le Jan-Watanabe connection we know that ∇A equals 0 on $\ker A(X_s(x))^\perp$, cf. Remark 2.3. ☐

Remark 3.10. i) Alternatively, one may rewrite the intrinsic local martingale by using $\ell_s = A(X_s(x))^*W_s\dot L_s$ to compute

$$\begin{aligned}&\nabla_v\left[(dF_s)_{X_s(x)}W_s\left(w+\int_0^s W_r^{-1}A(X_r(x))\ell_r\mathrm{d}r\right) - (F_s\circ X_s(x))\int_0^s\langle\ell_r,\mathrm{d}Z_r\rangle\right]\\&=\nabla dF_s\left(X_{s*}v, W_s\left(w+\int_0^s W_r^{-1}A(X_r(x))\ell_r\,\mathrm{d}r\right)\right)\\&\quad+(dF_s)_{X_s(x)}\nabla W_s\left(v, w+\int_0^s W_r^{-1}A(X_r(x))\ell_r\,\mathrm{d}r\right)\\&\quad+(dF_s)_{X_s(x)}W_s\int_0^s\nabla_{X_{r*}v}(W_r^{-1})A(X_r(x))\ell_r\,\mathrm{d}r\\&\quad+(dF_s)_{X_s(x)}W_s\int_0^s W_r^{-1}\nabla_{X_{r*}v}(A(X_r(x)))\ell_r\,\mathrm{d}r\\&\quad-(dF_s)_{X_s(x)}X_{s*}v\int_0^s\langle\ell_r,\mathrm{d}Z_r\rangle.\end{aligned}$$

Now substituting back to terms of L_s and using $\nabla_{X_{s*}v}(W_s^{-1})W_s = -W_s^{-1}\nabla_v W_s$ we obtain (3.26) by filtering out noise as in the proof above.

ii) A slightly different possibility to prove Theorem 3.9 is to start out by the local martingale $\nabla_v[(dF_s)_{X_s(x)}W_s w] = \nabla dF_s(X_{s*}v, W_s w) + (dF_s)_{X_s(x)}\nabla W_s(v,w)$ and take conditional expectation with respect to $\mathscr{F}_s^{X(x)}$ to assure that

$$\nabla dF_s(W_s v, W_s w) + (dF_s)_{X_s(x)}\widetilde{\nabla W}_s(v,w) \tag{3.27}$$

is a local martingale. As in part 3.1.3, we apply integration by parts to replace w by a $\big(\mathscr{F}_s^{X(x)}\big)$-adapted finite energy process L_s. The only difference is that one also has to filter out noise from equation (3.22) and write W_s instead of X_{s*} in the further computations.

iii) We have to admit, however, that we are not able to present a second order intrinsic local martingale comparable to (3.1), i.e. depending on two finite energy processes K_s and L_s. The reason for this failure is, that of course we can only filter terms with one linear factor containing noise, but there is no justification to compute e.g. $\mathbb{E}\big[b(TX_sK_s, TX_sL_s)|\mathscr{F}_s^{X(x)}\big] = b(W_sK_s, W_sL_s)$ for an arbitrary bilinear functional b on $TM \times TM$ since taking conditional expectation is just a linear operation.

So the only promising strategy to generalize the upper result would be to replace v by K_s and apply integration by parts which raises the same problems as described in Remark 3.8.

3.3. Hessian formulae containing curvature expressions

According to an observation by Marc Arnaudon and Anton Thalmaier, the quite formal process $(\widetilde{\nabla W}_s)$ of the preceding paragraph can be translated into terms of the Riemannian curvature tensor on M the following way. Note that $\mathrm{d}^\nabla X_s$ stands for stochastic integration with respect to the Itô integral.

Lemma 3.11. *The process* $(\widetilde{\nabla W}_s)$ *is given by* $\widetilde{\nabla W}_0 = 0$ *and*

$$W_s\mathrm{d}(W_s^{-1}\widetilde{\nabla W}_s) = R(\mathrm{d}^\nabla X_s, W_s)W_s - \frac{1}{2}\nabla\,\mathrm{Ric}^\sharp(W_s, W_s)\mathrm{d}s - \frac{1}{2}d^*R(W_s)W_s\mathrm{d}s, \tag{3.28}$$

where $\nabla\,\mathrm{Ric}^\sharp(v,w) = \nabla_v\,\mathrm{trace}\,R(w,\,\cdot\,)\,\cdot$ *and* $d^*R(w) = -\,\mathrm{trace}\,\nabla\,.\,R(\,\cdot\,,w)$.

Proof. Equation (3.28) is an application of a more general theorem about commutation of derivatives with respect to space and time of a C^1-family of semimartingales taking values in some vector bundle over M. This result is given in [A-Th 4], Theorem 4.5, which tells in our case that

$$\begin{aligned} D\nabla W_s &= \nabla_{TX_s}DW_s + R(\mathrm{d}^\nabla X_s, TX_s)W_s + R(\mathrm{d}X_s, TX_s)DW_s \\ &\quad + \frac{1}{2}\nabla R(\mathrm{d}X_s, \mathrm{d}X_s, TX_s)W_s - \frac{1}{2}R(DTX_s, \mathrm{d}X_s)W_s \end{aligned} \tag{3.29}$$

(recall that the curvature tensor R is antisymmetric in the first two arguments).

Because of W_s satisfying a pathwise equation and (3.24), the third and last term on the right hand side both equal 0. Moreover, the defining stochastic differential equation for W_s (in the case of $V = 0$) reads

$$DW_s = -\frac{1}{2}\,\mathrm{Ric}^\sharp(W_s)\mathrm{d}s = -\frac{1}{2}R(W_s, \mathrm{d}X_s)\mathrm{d}X_s,$$

which we covariantly differentiate by ∇ (in direction $TX_s(\,\cdot\,)$, since we are tracking paths of $X_s(x)$) and cancel out the last two of the four occuring terms by using (3.25). So our preliminary result is

$$\begin{aligned} D\nabla W_s = &-\frac{1}{2}\nabla R(TX_s, W_s, \mathrm{d}X_s)\mathrm{d}X_s - \frac{1}{2}R(\nabla W_s, \mathrm{d}X_s)\mathrm{d}X_s \\ &+ R(\mathrm{d}^\nabla X_s, TX_s)W_s + \frac{1}{2}\nabla R(\mathrm{d}X_s, \mathrm{d}X_s, TX_s)W_s. \end{aligned} \tag{3.30}$$

As our goal is an expression for $\widetilde{\nabla W}_s$, we write

$$\begin{aligned} W_s\mathrm{d}(W_s^{-1}\nabla W_s) &= W_s\mathrm{d}(W_s^{-1}//_{0,s}//_{0,s}^{-1}\nabla W_s) \\ &= W_s\mathrm{d}(W_s^{-1}//_{0,s})//_{0,s}^{-1}\nabla W_s + //_{0,s}\mathrm{d}(//_{0,s}^{-1}\nabla W_s) \\ &= \frac{1}{2}R(\nabla W_s, \mathrm{d}X_s)\mathrm{d}X_s + D\nabla W_s, \end{aligned} \tag{3.31}$$

where for the last step we have used (2.10) to compute

$$\begin{aligned} \mathrm{d}(W_s^{-1}//_{0,s}) = \mathrm{d}Q_s^{-1} &= \frac{\partial}{\partial s}(Q_s^{-1})\,\mathrm{d}s = -Q_s^{-1}\dot{Q}_sQ_s^{-1}\mathrm{d}s \\ &= \frac{1}{2}Q_s^{-1}(\mathrm{Ric}_{//_{0,s}}\,Q_s)Q_s^{-1}\mathrm{d}s = \frac{1}{2}W_s^{-1}(\mathrm{Ric}^\sharp_{X_s(x)}\,W_s)W_s^{-1}//_{0,s}. \end{aligned}$$

By putting (3.30) into (3.31) we obtain

$$\begin{aligned} W_s\mathrm{d}(W_s^{-1}\nabla W_s) = &-\frac{1}{2}\nabla R(TX_s, W_s, \mathrm{d}X_s)\mathrm{d}X_s + R(\mathrm{d}^\nabla X_s, TX_s)W_s \\ &+\frac{1}{2}\nabla R(\mathrm{d}X_s, \mathrm{d}X_s, TX_s)W_s. \end{aligned} \tag{3.32}$$

Filtering out noise turns TX_s into W_s in every term on the right hand side and using the abbreviations defined in the lemma we end up with (3.28). $\square$

At this point we are given all prerequisites to prove the following main result.

Theorem 3.12 (Intrinsic Hessian representation formula).
We assume that we are given a space-time-harmonic function F on M, where $L = \frac{1}{2}\Delta$ and thus X is a Brownian motion on M. In addition to this, we fix $t > 0$ and stopping times $0 < \sigma < \tau$ chosen just as in Theorem 3.4, which means $\sigma = \tilde{\sigma} \wedge \frac{t}{2}$ and $\tau = \tilde{\tau} \wedge t$; $0 < \tilde{\sigma} < \tilde{\tau}$ the first exit times of two open and relatively compact domains satisfying $x \in D_1 \subset \bar{D}_1 \subset D_2$.
Now if ∇ denotes the Le Jan-Watanabe connection on M, we have

$$\begin{aligned} &\mathrm{Hess}_x F_0(v,w) \equiv (\nabla dF_0)(v,w) \\ &\quad= \mathbb{E}\Bigg[F_\sigma \circ X_\sigma(x)\bigg\{-\int_0^\sigma \Big\langle W_r \int_0^r W_s^{-1} R\left(//_{0,s}\mathrm{d}B_s, W_s v\right) W_s\dot{L}_r, //_{0,r}\mathrm{d}B_r\Big\rangle \\ &\qquad\qquad + \frac{1}{2}\int_0^\sigma \Big\langle W_r \int_0^r W_s^{-1}\left(\nabla\,\mathrm{Ric}^\sharp + \mathrm{d}^* R\right)\left(W_s v, W_s\dot{L}_r\right)\mathrm{d}s, //_{0,r}\mathrm{d}B_r\Big\rangle\bigg\}\Bigg] \\ &\qquad + \mathbb{E}\left[F_\tau \circ X_\tau(x)\int_0^\sigma \Big\langle W_r\dot{L}_r, //_{0,r}\mathrm{d}B_r\Big\rangle \int_\sigma^\tau \Big\langle W_r\dot{K}_r, //_{0,r}\mathrm{d}B_r\Big\rangle\right] \end{aligned} \tag{3.33}$$

where K, L are bounded $\big(\mathscr{F}_s^{X(x)}\big)$-adapted finite energy processes, more exactly $\int_0^t \|\dot K_s\|^2 \mathrm{d}s$ and $\int_0^t \|\dot L_s\|^2 \mathrm{d}s$ are both contained in $L^{1+\alpha}(\mathbb{P})$ for some $\alpha > 0$), such that $K_s = v$ for $s \le \sigma$ and $K_s = 0$ for $s \ge \tau$, respectively, $L_0 = w$ and $L_s = 0$ for $s \ge \sigma$.
As in the intrinsic first order formula, B_s herein is an n-dimensional T_xM-valued Brownian motion that is given by $//_{0,s}\mathrm{d}B_s = A(X_s(x))\mathrm{d}Z_s = \mathrm{d}^\nabla X_s$.

Proof. The intrinsic integration by parts formula applied to the local martingale of Lemma 3.9 provides that

$$\begin{aligned}
&\nabla dF_s(W_s v, W_s L_s) + (dF_s)_{X_s(x)} \widetilde{\nabla W}_s(v, L_s) \\
&\quad - (dF_s)_{X_s(x)} W_s \int_0^s W_r^{-1} \widetilde{\nabla W}_r(v, \dot L_r)\mathrm{d}r \\
&\quad - (dF_s)_{X_s(x)} W_s v \int_0^s \langle W_r \dot L_r, A(X_r(x))\mathrm{d}Z_r\rangle \\
&\stackrel{m}{=} \nabla dF_s(W_s v, W_s L_s) + (dF_s)_{X_s(x)} \widetilde{\nabla W}_s(v, L_s) \\
&\quad - F_s \circ X_s(x) \int_0^s \langle \widetilde{\nabla W}_r(v, \dot L_r), A(X_r(x))\mathrm{d}Z_r\rangle \\
&\quad - (dF_s)_{X_s(x)} W_s v \int_0^s \langle W_r \dot L_r, A(X_r(x))\mathrm{d}Z_r\rangle.
\end{aligned}$$

Reasoning as the proof of Theorem 3.4 shows that the right hand side of the last equation is a true martingale on $[0, \tau]$, proven that $\sup_{s\le\tau} \|\widetilde{\nabla W}_s(v, \cdot)\|$ is contained in $L^p(\mathbb{P})$ for some $p > 1$.
But the other hand, by Lemma 3.11 we have

$$\widetilde{\nabla W}_s = W_s \int_0^s W_r^{-1} \left[R(//_{0,r}\mathrm{d}B_r, W_r)W_r - \frac{1}{2}\Big(\nabla \operatorname{Ric}^\sharp(W_r, W_r) + d^*R(W_r)W_r\Big)\,\mathrm{d}r \right].$$

As we can see from Lemma 4.17 below, $\sup_{s\le\tau}\|W_s\|$ and $\sup_{s\le\tau}\|W_s^{-1}\|$ are finite due to Ricci bounds on the compact neighbourhood $\bar D_2$ of x (in which $(X_s(x))_{s\le\tau}$ takes its values). Using this and the boundedness of the involved multilinear curvature terms on $\bar D_2$, a Burkholder-Gundy argument provides the necessary integrability of $\sup_{s\le\tau}\|\widetilde{\nabla W}_s(v, \cdot)\|$. So taking expectations at times 0 and τ yields

$$\begin{aligned}
&(\nabla dF_0)(v, w) \\
&= \mathbb{E}\left[F_\tau \circ X_\tau(x) \left\{ -\int_0^\sigma \left\langle W_r \int_0^r W_s^{-1} R\left(//_{0,s}\mathrm{d}B_s, W_s v\right) W_s \dot L_r, //_{0,r}\mathrm{d}B_r \right\rangle \right.\right. \\
&\qquad \left.\left. + \frac{1}{2}\int_0^\sigma \left\langle W_r \int_0^r W_s^{-1}\left(\nabla \operatorname{Ric}^\sharp + \mathrm{d}^*R\right)\left((W_s v, W_s \dot L_r)\right)\mathrm{d}s, //_{0,r}\mathrm{d}B_r \right\rangle \right\}\right] \\
&\quad - \mathbb{E}\left[(dF_\tau)_{X_\tau(x)} W_\tau v \int_0^\sigma \langle W_r \dot L_r, A(X_r(x))\mathrm{d}Z_r\rangle \right].
\end{aligned}$$

In the first term we additionally take conditional expectation with respect to $\mathscr{F}_\sigma^{X(x)}$ to turn $F_\tau \circ X_\tau(x)$ into $F_\sigma \circ X_\sigma(x)$, and in the last expectation the same argument as in Remark 3.7 (using the intrinsic versions of O_s^1 and O_s^2 there) finally provides (3.33). □

Corollary 3.13. *In the special case of $F_s = P_{t-s}f$ denoting the minimal heat semigroup on M with respect to $L = \frac{1}{2}\Delta$, $f \in L^\infty(M)$, we derive the stochastic representation formula*

$$\begin{aligned}
&\operatorname{Hess}_x P_t f(v,w) \\
&= \mathbb{E}\Bigg[f(X_t(x))1_{t<\zeta(x)}\Bigg\{-\int_0^\sigma \left\langle W_r \int_0^r W_s^{-1} R\left(//_{0,s}\mathrm{d}B_s, W_s v\right) W_s \dot{L}_r, //_{0,r}\mathrm{d}B_r\right\rangle \\
&\qquad + \frac{1}{2}\int_0^\sigma \left\langle W_r \int_0^r W_s^{-1}\left(\nabla \operatorname{Ric}^\sharp + \mathrm{d}^* R\right)\left(W_s v, W_s \dot{L}_r\right)\mathrm{d}s, //_{0,r}\mathrm{d}B_r\right\rangle \\
&\qquad + \int_0^\sigma \left\langle W_r \dot{L}_r, //_{0,r}\mathrm{d}B_r\right\rangle \int_\sigma^\tau \left\langle W_r \dot{K}_r, //_{0,r}\mathrm{d}B_r\right\rangle\Bigg\}\Bigg],
\end{aligned} \tag{3.34}$$

K and L satisfying the same assumptions as in the theorem above.

Proof. This is an immediate consequence of the strong Markov property, which here reads

$$P_{t-\tau}f(X_\tau(x)) = \mathbb{E}\left[F_t \circ X_t(x)1_{\{t<\zeta(x)\}} | \mathscr{F}_\tau^{X(x)}\right],$$

and the approximation argument that we have already used in the proof of Theorems 2.19 and 3.4, so we refer to that part for details. □

The result of Theorem 3.12 may be used to obtain the following Bismut formula for $\operatorname{Hess}_x \log p(t,\,\cdot\,,y)$.

Corollary 3.14 (Bismut formula for the Hessian of the logarithmic heat kernel). *For $p(t,x,y)$ given as in (2.28) (related to $L = \frac{1}{2}\Delta$) we have the stochastic representation ($x \in M$, $v, w \in T_x M$)*

$$\begin{aligned}
\operatorname{Hess}_x(\log p(t,\,\cdot\,,y))(v,w) &= \mathbb{E}\left[J_\sigma + I_\sigma^1 I_\tau^2 | X_t(x) = y\right] \\
&\quad - \mathbb{E}\left[I_\sigma^1 | X_t(x) = y\right]\ \mathbb{E}\left[I_\tau^2 | X_t(x) = y\right],
\end{aligned} \tag{3.35}$$

where I^1, I^2, J are given by

$$\begin{aligned}
J_\sigma &:= -\int_0^\sigma \left\langle W_r \int_0^r W_s^{-1} R\left(//_{0,s}\mathrm{d}B_s, W_s v\right) W_s \dot{L}_r, //_{0,r}\mathrm{d}B_r\right\rangle \\
&\qquad + \frac{1}{2}\int_0^\sigma \left\langle W_r \int_0^r W_s^{-1}\left(\nabla \operatorname{Ric}^\sharp + \mathrm{d}^* R\right)\left(W_s v, W_s \dot{L}_r\right)\mathrm{d}s, //_{0,r}\mathrm{d}B_r\right\rangle, \\
I_\sigma^1 &:= \int_0^\sigma \left\langle W_r \dot{L}_r, //_{0,r}\mathrm{d}B_r\right\rangle, \qquad \textit{and} \qquad I_\tau^2 := \int_\sigma^\tau \left\langle W_r \dot{K}_r, //_{0,r}\mathrm{d}B_r\right\rangle,
\end{aligned} \tag{3.36}$$

σ, τ, K, L defined as in Theorem (3.12) above.
In particular, if M is flat, i.e. $R \equiv 0$, one has the representation of the Hessian as a covariance of two real martingales taken with respect to the Brownian bridge measure $\mathbb{P}_{x,y}^t$

$$\operatorname{Hess}_x(\log p(t,\,\cdot\,,y))(v,w) = \operatorname{Cov}^{\mathbb{P}_{x,y}^t}\left(\int_0^\sigma \left\langle \dot{L}_r, \mathrm{d}B_r\right\rangle \int_\sigma^\tau \left\langle \dot{K}_r, \mathrm{d}B_r\right\rangle\right).$$

Proof. As in Corollary 2.21, we have to compare for $f \in C_c(M)$ the equations

$$\operatorname{Hess}_x P_t f(v,w) = \int_M \operatorname{Hess}_x(p(t,\,\cdot\,,y))(v,w) f(y)\operatorname{vol}(\mathrm{d}y) \tag{3.37}$$

and (3.33), which reads

$$\begin{aligned}\operatorname{Hess}_x P_t f(v,w) &= \mathbb{E}\left[f \circ X_t(x)\left(J_\sigma + I^1_\sigma I^2_\tau\right)\right] \\ &= \int_M \mathbb{E}\left[\left(J_\sigma + I^1_\sigma I^2_\tau\right) | X_t(x) = y\right] p(t,x,y) f(y)\operatorname{vol}(\mathrm{d}y).\end{aligned} \tag{3.38}$$

This implies $\operatorname{Hess}_x(p(t,\,\cdot\,,y))(v,w)/p(t,x,y) = \mathbb{E}\left[\left(J_\sigma + I^1_\sigma I^2_\tau\right)|X_t(x) = y\right]$. On the other hand, the chain rule for $g \in C^2(M)$ provides

$$\begin{aligned}\operatorname{Hess}_x(\log g)(v,w) &= \frac{\operatorname{Hess}_x g(v,w)}{g(x)} - \frac{\langle \operatorname{grad}_x g, v\rangle\langle \operatorname{grad}_x g, w\rangle}{g^2(x)} \\ &= \frac{\operatorname{Hess}_x g(v,w)}{g(x)} - \langle \operatorname{grad}_x \log g, v\rangle\langle \operatorname{grad}_x \log g, w\rangle.\end{aligned} \tag{3.39}$$

By $g = p(t,\,\cdot\,,y)$ and twice using (2.39) (with I^1 and I^2 instead of I respectively) we are finished.
In the flat case, all curvature terms vanish and therefore $J \equiv 0$. Moreover, the damped transport then fulfills $DW_s = 0$, $W_0 = \mathrm{id}_{T_xM}$, and hence $W_s = //_{0,s}$. □

Another special case of the Hessian representation formula is derived when M is Ricci parallel, i.e. $\nabla \operatorname{Ric} = 0$. This assumption causes that $d^*R = 0$ as well, a fact which the author was indicated to by Marc Arnaudon, who also communicated the following short proof.

Lemma 3.15. *Let (M,g) be a Riemannian manifold of dimension n, ∇ the Levi-Civita connection on M and $x \in M$. Then $\nabla \operatorname{Ric}(x) = 0$ implies that also $d^*R(x) = 0$.*

Proof. Fix an orthonormal frame (e_i) in a neighbourhood of x, such that at x we have $\nabla_{e_i} e_j = 0$ for all i,j. We compute by using the Bianchi identity for the second equality and the symmetry of the curvature tensor for the sixth equality, that in T_xM

$$\begin{aligned}\nabla \operatorname{Ric}^\sharp(e_k,e_l) - \nabla \operatorname{Ric}^\sharp(e_l,e_k) &= \sum_{i=1}^n \left(\nabla_{e_k} R(e_l,e_i)e_i - \nabla_{e_l} R(e_k,e_i)e_i\right) \\ &= -\sum_{i=1}^n \nabla_{e_i} R(e_k,e_l)e_i \\ &= -\sum_{j=1}^n\sum_{i=1}^n \langle \nabla_{e_i} R(e_k,e_l)e_i, e_j\rangle\, e_j \\ &= -\sum_{j=1}^n\sum_{i=1}^n \langle \nabla_{e_i}\left(R(e_k,e_l)e_i\right), e_j\rangle\, e_j \\ &= -\sum_{j=1}^n\sum_{i=1}^n e_i\,\langle R(e_k,e_l)e_i, e_j\rangle\, e_j\end{aligned}$$

$$
\begin{aligned}
&= -\sum_{j=1}^{n}\sum_{i=1}^{n} e_i \left\langle R(e_j,e_i)e_l, e_k\right\rangle e_j \\
&= -\sum_{j=1}^{n}\sum_{i=1}^{n} \left\langle \nabla_{e_i}\left(R(e_j,e_i)e_l\right), e_k\right\rangle e_j \\
&= -\sum_{j=1}^{n}\sum_{i=1}^{n} \left\langle \nabla_{e_i} R(e_j,e_i)e_l, e_k\right\rangle e_j \\
&= -\sum_{j=1}^{n} \left\langle d^* R(e_j)e_l, e_k\right\rangle e_j
\end{aligned}
$$

(the notation $e_i\langle R(e_j,e_i)e_l,e_k\rangle$ means the action of the – locally well-defined – vector field e_i on the smooth function given by the inner product).
For this reason the assumption $\nabla \operatorname{Ric} = 0$ and thus $\nabla \operatorname{Ric}^\sharp = 0$ yields

$$
\sum_{j=1}^{n} \left\langle d^* R(e_j)e_l, e_k\right\rangle e_j = 0, \qquad 1 \le k,l \le n
$$

and consequently

$$
\left\langle d^* R(e_j)e_l, e_k\right\rangle = 0
$$

for all j,k,l. This proves $d^*R(x) = 0$. □

Corollary 3.16. *If in addition to the situation of Theorem 3.12 our manifold M is Ricci parallel, which means $\nabla \operatorname{Ric} = 0$ on M, we have the stochastic Hessian representation formula*

$$
\begin{aligned}
&\operatorname{Hess}_x F_0(v,w) \\
&= -\mathbb{E}\left[F_\sigma \circ X_\sigma(x) \int_0^\sigma \left\langle E_r \int_0^r E_{-s}\, //_{0,s}^{-1} R\left(//_{0,s}\mathrm{d}B_s, //_{0,s}E_s v\right) //_{0,s} E_s \dot L_r, \mathrm{d}B_r\right\rangle\right] \\
&\quad + \mathbb{E}\left[F_\tau \circ X_\tau(x) \int_0^\sigma \left\langle E_r \dot L_r, \mathrm{d}B_r\right\rangle \int_\sigma^\tau \left\langle E_r \dot K_r, \mathrm{d}B_r\right\rangle\right].
\end{aligned}
\tag{3.40}
$$

where $(E_r)_{r\in\mathbb{R}}$ denotes the family of (deterministic) endomorphisms on T_xM given by

$$
E_r := e^{-\frac{r}{2}\operatorname{Ric}_x^\sharp}.
$$

Proof. According to Lemma 3.15 the terms of equation (3.33) depending linearly on $\nabla \operatorname{Ric}^\sharp$ and d^*R vanish here. Hence it remains to show that

$$
W_r = //_{0,r} E_r.
$$

But $W_r = //_{0,r} Q_r$ with the (stochastic) endomorphism Q_r on T_xM defined by (2.10). The assumption $\nabla \operatorname{Ric} = 0$ implies $\operatorname{Ric}^\sharp_{X_r(x)} = //_{0,r} \operatorname{Ric}^\sharp_x //_{0,r}^{-1}$, and consequently $\operatorname{Ric}_{//_{0,r}} = \operatorname{Ric}^\sharp_x$ in the defining equation for Q_s. For this reason we find that Q_r solves the deterministic linear differential equation with constant coefficient

$$
\dot Q_r = -\frac{1}{2}\operatorname{Ric}^\sharp_x Q_r, \qquad Q_0 = \mathrm{id}_{T_xM}.
$$

Obviously the unique solution to this is $Q_r = e^{-\frac{r}{2}\operatorname{Ric}^\sharp_x}$. □

Chapter 4

Gradient and Hessian estimates

A suitable field for deterministic applications of our stochastic representation formulae derived in the previous chapters are gradient and Hessian estimates of harmonic or space-time-harmonic functions on Riemannian manifolds.
However, it takes some effort to prove good estimates from our stochastic representations of the derivatives. This fact is mainly based on two reasons. First, in the main theorems there is still a considerable degree of freedom in the choice of suitable finite energy processes. And the second and more difficult question is how to separate the factors of the diffusion put in the (space-time) harmonic function and those consisting of stochastic integrals with respect to the driving diffusion, of which the expectation is taken at last.

For the sake of brevity, we will write $|\cdot|$ for the norm on TM (mainly on T_xM) throughout the whole chapter.

In particular and as the easiest example, take the case of a (relatively compact) regular domain $D \subset M$ and some interior point $x \in D$. For the gradient at x of a harmonic function $u \in C^2(D)$, we have proven

$$\langle \operatorname{grad}_x u, v\rangle = \mathbb{E}\left[u(X_{\tau\wedge t}(x)) \int_0^{\tau\wedge t} \langle W_r \dot{K}_r, //_{0,r} \mathrm{d}B_r\rangle\right], \tag{4.1}$$

where $t > 0$, τ the first exit time of $X_r(x)$ from D, $v \in T_xM$, $(K_r)_{r\in[0,t]}$ a suitable finite energy process with $K_0 = v$, $K_t = 0$, and $X_r(x)$ a Brownian motion on M starting in x (such that $\mathrm{d}X_r = //_{0,r}\mathrm{d}B_r$).
For the stochastic integral in (4.1), one can deduce from results due to Thalmaier and Wang, cf. [Th-W], that for $|K_0| = |v| = 1$ and $r(x) := \operatorname{dist}(x, \partial D)$

$$\left\| \int_0^{\tau\wedge t} \langle W_r \dot{K}_r, //_{0,r} \mathrm{d}B_r\rangle \right\|_{L^2(\mathbb{P})} \leq C(r(x)) := \sqrt{c_1 r^{-2}(x) + c_2 r^{-1}(x) + c_3}$$

with constants c_1, c_2, c_3 depending on n and a lower Ricci bound on D.
Hence by passing over to the L^∞-norm of $u|_D$ in (4.1), we immediately derive an estimate of the form

$$|\operatorname{grad}_x u| \leq \|u\|_D \sqrt{c_1 r^{-2}(x) + c_2 r^{-1}(x) + c_3}. \tag{4.2}$$

Obviously, this method transfers to an estimate for the gradient of heat semigroups.

For further applications one could consider the case of positive harmonic functions on M.

An analytic argument due to Cheng and Yau ([Ch-Y]) provides a local estimate for strictly positive harmonic functions u of the form

$$|\operatorname{grad}_x(\log u)| = \frac{|\operatorname{grad}_x u|}{u(x)} \le c(n)\frac{1+kr(x)}{r(x)}, \tag{4.3}$$

where $r(x) := \operatorname{dist}(x, \partial D)$. $c(n)$ is a positive constant depending on the dimension n of M and $k \ge 0$ is chosen such that $\inf_{y\in D,\, w\in T_yM,\, |w|=1} \operatorname{Ric}(w,w) \ge -(n-1)k^2$ (or briefly: $\operatorname{Ric} \ge -(n-1)k^2$ on D). This theorem is also formulated and proven in [Sch], cf. Theorem 1.1.

Observe that the order in $r(x)$ on the right hand side of (4.3) corresponds to that of $C(r(x))$ above.

However, it is not clear how to derive an estimate of the same type by means of (4.1). Although we know that $u(X_r(x))$ is a positive martingale (since $\bar{D}$ is compact), and consequently $\mathbb{E}\left[|u(X_{\tau\wedge r}(x))|\right] = u(x)$ for all $r \ge 0$, it is not possible to split the expectation of the product on the right hand side of (4.1) by Hölder's inequality in order to obtain an estimate of the form $|\operatorname{grad}_x u| \le Cu(x)$. The reason for this failure is that (in general) the stochastic integral with respect to B_s has infinite L^∞-norm. On the other hand, for any $p > 1$ the expectation of $u(X_s(x))^p$ increases in time due to the Itô formula.

As a first step towards similar results for positive harmonic functions by means of (2.37), in [Th-W] the authors use the Cauchy-Schwartz inequality in connection with

$$\|u(X_s(x))\|^2_{L^2(\mathbb{P})} \equiv \mathbb{E}[u^2(X_s(x))] \le \mathbb{E}[u(X_s(x))]\,\|u\|_D = u(x)\|u\|_D$$

to obtain the local estimate

$$|\operatorname{grad}_x u| \le \sqrt{u(x)\|u\|_D}\; C(r(x)).$$

We will present a thorough summary of the results in [Th-W] for the gradient case, which particularly consists of estimating the stochastic integral of the representation formula, with the intention to transfer them for similar Hessian estimates by applying our result of Theorem 3.12.

Returning to pointwise formulae (in the sense that the right hand side merely depends on $u(x)$), we conclude this chapter by focusing on manifolds, which are rotationally invariant with respect to the point the gradient or Hessian is taken at. In this particular case (and given that $L = \frac{1}{2}\Delta$), there is a certain independence property that allows us to split the expectation on the right hand side of (4.1) into $u(x)$ and the 1-norm of the stochastic integral.

4.1. Local gradient estimate

We reproduce the results of [Th-W], §4-6. A few lemmas are slightly generalized, since we need them for the Hessian results afterwards.

For the rest of the chapter we assume that (M, g) is a Riemannian manifold of dimension $n \ge 2$.

$X_s(x)$ shall denote a diffusion starting in $x \in M$ with respect to the (elliptic) generator $L = \frac{1}{2}\Delta_M + V$, V an arbitrary vector field. As we explained in the previous chapters, we may assume that the coefficients A and A_0 of the Stratonovich equation defining X are related to an isometric embedding of (M, g) in Euclidean space, cf. Remark 1.1 ii). Then the Levi-Civita and the Le Jan-Watanabe connections on TM coincide (cf. Remark 2.4) and consequently we are able to apply the intrinsic gradient formula to derive results about diffusions with generator $L = \frac{1}{2}\Delta_M + V$, Δ_M the Laplace-Beltrami operator on M.

We formulate the main theorem of this paragraph depending on a domain D and a C^2-function f on D. Explicit estimates then are derived by particular choices of f and D.

Theorem 4.1. *For $x \in M$ let D denote an open regular domain containing x, i.e. a relatively compact neighbourhood of x with smooth boundary (which may be empty, e.g. $D = M$ in the case of compact M).*
Consider $f \in C^2(\bar{D})$ with $0 < f|_D \leq 1$ and $f|_{\partial D} = 0$ and write τ for the first exit time of $X_s(x)$ from D.
In addition to this, let $t > 0$ and $F : [0, t] \times M \to \mathbb{R}$ be space-time-harmonic on $[0, t] \times \bar{D}$ (in particular, continuous on $[0, t] \times \bar{D}$ with jointly continuous first and second derivatives on $[0, t[$ times a neighbourhood of $\bar{D}$). Then we have

$$|\operatorname{grad}_x F_0(\,\cdot\,)| \leq \frac{\sqrt{c(f)}}{f(x)\sqrt{1 - e^{-c(f)t}}} \left(\|F_t\|_D \vee \sup_{[0,t]\times\partial D} |F| \right) \tag{4.4}$$

with the convention that coefficient on the right hand side equals $\frac{1}{f(x)\sqrt{t}}$ if $c(f) = 0$.
The constant $c(f)$ depending on f and a lower curvature bound on D is given by

$$c(f) := \sup_D \left(-f^2 k_V + 3|\operatorname{grad} f|^2 - 2f \left(\frac{1}{2}\Delta + V \right) f \right)_+ \tag{4.5}$$

(the index $+$ denotes the positive part), where

$$k_V := \inf_{y\in D,\, w\in T_yM,\, |w|=1} (\operatorname{Ric}(w, w) - 2\langle \nabla_w V, w\rangle). \tag{4.6}$$

Proof. As an immediate consequence of our intrinsic representation formula (2.37), for $v \in T_xM$ holds

$$\begin{aligned}
|\langle \operatorname{grad}_x F_0(\,\cdot\,), v\rangle| &= \left| \mathbb{E}\left[F_{\tau\wedge t}(X_{\tau\wedge t}(x)) \int_0^{\tau\wedge t} \langle W_r \dot{K}_r, //_{0,r} \mathrm{d}B_r\rangle \right] \right| \\
&\leq \|F_{\tau\wedge t}(X_{\tau\wedge t}(x))\|_{L^2(\mathbb{P})} \left\| \int_0^{\tau\wedge t} \langle W_r \dot{K}_r, //_{0,r} \mathrm{d}B_r\rangle \right\|_{L^2(\mathbb{P})} \\
&\leq \left(\|F_t\|_D \vee \sup_{[0,t]\times\partial D} |F| \right) \left\| \int_0^{\tau\wedge t} \langle W_r \dot{K}_r, //_{0,r} \mathrm{d}B_r\rangle \right\|_{L^2(\mathbb{P})}
\end{aligned}$$

due to $X_{\tau\wedge t}(x) \in \partial D$ on $\{\tau \leq t\}$ and $X_t(x) \in D$ on $\{t < \tau\}$.
Of course, we could directly estimate $|F_{\tau\wedge t}(X_{\tau\wedge t}(x))| \leq \|F_t\|_D \vee \sup_{[0,t]\times\partial D} |F|$ to obtain the L^1-norm on the right hand side. However, we give modifications in the positive case below, where the L^2-norm of $F_{\tau\wedge t}(X_{\tau\wedge t}(x))$ has some advantage over merely using the supremum, and consideration of the L^1-norm of the stochastic integral does not yield essentially better bounds either.

According to the notions of Theorem 2.27, K_r above denotes an arbitrary bounded finite energy process with $K_0 = v$ and $K_{\tau\wedge t} = 0$. So the theorem is proven if for any given $v \in T_xM$ with $|v| = 1$ we are able to construct such a K_r which fulfills

$$\left\| \int_0^{\tau\wedge t} \langle W_r \dot{K}_r, //_{0,r} \mathrm{d}B_r \rangle \right\|_{L^2(\mathbb{P})} \le \frac{c(f)}{f^2(x)(1 - e^{-c(f)t})} .$$

This will be carried out in several steps below. □

Remark 4.2. Of course the estimate

$$\|F_{\tau\wedge t}(X_{\tau\wedge t}(x))\|_{L^2(\mathbb{P})} \le \|F_t\|_D \vee \sup_{[0,t]\times\partial D} |F|$$

needs to be interpreted and specialized with respect to the particular examples for F we have in mind as there are:

i) In the case of $F_s = P_{t-s}g$ denoting the heat semigroup on M for $g \in L^\infty(M)$, we have $|P_{t-s}g(y)| \le \|g\|_\infty$ for all $0 \le s \le t$ and $y \in M$. Hence the theorem modifies to

$$|\operatorname{grad}_x P_t g(\,\cdot\,)| \le \frac{\sqrt{c(f)}}{f(x)\sqrt{1 - e^{-c(f)t}}} \|g\|_\infty \tag{4.7}$$

Of course, this is a direct consequence of (2.38) as well.

ii) If we consider the heat semigroup acting on positive bounded measurable functions, we can use the fact that P_s preserves positivity and the martingale property of $P_{t-s}g(X_s(x))$ to estimate the L^2-norm by

$$\begin{aligned} \|P_{t-\tau\wedge t}g(X_{\tau\wedge t}(x))\|_{L^2(\mathbb{P})} &= \left(\mathbb{E}\left[g(X_t(x))1_{\{t<\zeta(x)\}}\right]^2\right)^{1/2} \\ &\le \left(\|g\|_\infty \mathbb{E}\left[g(X_t(x))1_{\{t<\zeta(x)\}}\right]\right)^{1/2} = \sqrt{\|g\|_\infty P_t g(x)}. \end{aligned}$$

so the estimate reads

$$|\operatorname{grad}_x P_t g(\,\cdot\,)| \le \frac{\sqrt{c(f)}}{f(x)\sqrt{1 - e^{-c(f)t}}} \sqrt{P_t g(x)\|g\|_\infty}. \tag{4.8}$$

iii) For the case of a L-harmonic function u on D that is continuous on $\bar{D}$, we have $\|u(X_{\tau\wedge t}(x))\|_{L^2(\mathbb{P})} \le \|u\|_D$ and pass over to the limit $t \to \infty$ in the estimate which lets the root in the denominator tend to 1. Thus

$$|\operatorname{grad}_x u(\,\cdot\,)| \le \frac{\sqrt{c(f)}}{f(x)} \|u\|_D. \tag{4.9}$$

iv) Combination of the last two cases yields:
If u is positive L-harmonic on D and continuous on $\bar{D}$, then

$$|\operatorname{grad}_x u(\,\cdot\,)| \le \frac{\sqrt{c(f)}}{f(x)} \sqrt{u(x)\|u\|_D}. \tag{4.10}$$

To procede with the proof of Theorem 4.1, we construct K_r in terms of the C^2-function f on $\bar{D}$ similarly to Part 2.3.
The basic idea for this procedure is as follows: our process K_r has to decay (path by path) from the initial value v to 0 within time $\tau \wedge t$ continuously and differentiable with respect to time at a.e. $r \in [0, \tau \wedge t]$ (since it has absolutely continuous paths). This is no problem for those paths of $X_r(x)$, where $t \le \tau$, i.e. the paths that do not reach ∂D within time t. Thus in the compact case $M = D$, we could choose $K_r := (t - (r \wedge t))\frac{v}{t}$.
However, due to the ellipticity of L, in general $P\{\tau < \varepsilon\} > 0$ for $\varepsilon > 0$, i.e. with positive probability we have paths of $X_r(x)$ where K_r has to turn to 0 within $[0, \varepsilon]$.
For this reason, we use f to introduce a time scale such that all paths of $X_{\bullet}(x)$ stay in D forever, and hence define the following time-change processes.

Definition 4.3. For $f \in C^2(\bar{D})$, $0 < f \le 1$ on D, $f|_{\partial D} = 0$, let

$$T(r) := \int_0^r f^{-2}(X_s(x))\mathrm{d}s, \qquad 0 \le r \le \tau, \tag{4.11}$$

and

$$\tau(r) := \inf\{s \ge 0 : T(s) \ge r\}, \qquad r \in \overline{\mathbb{R}}_+. \tag{4.12}$$

We will see below that $\tau(\infty) = \tau$ a.s. and $\tau(r) < \tau$ for any finite r.
Thus let $X'_r(x) := X_{\tau(r)}(x)$ denote the time-changed diffusion on D, adapted to $\big(\mathscr{F}^{X(x)}_{\tau(r)}\big)$.

Remark 4.4. Obviously, these processes are continuous, increasing and inverse to each other, namely $\tau(T(r)) = r$ for $r \in [0, \tau]$ and $T(\tau(r)) = r$ for arbitrary r. We will show in a moment that $T(\tau) = \infty$ a.s. (and $T(\tau)$ exactly is the lifetime $\zeta'(x)$ of $X'_r(x)$). This fact and $T(\tau(t)) = t$ imply $\tau(t) < \tau$. Moreover, $f \le 1$ yields $f^{-2} \ge 1$, thus $T(\tau(t)) \ge \tau(t)$ and consequently $\tau(t) \le t$. So finally we arrive at

$$\tau(t) \le \tau \wedge t.$$

Lemma 4.5. *$X'_r(x)$ is a diffusion starting in x with respect to the generator $L' := f^2 L = f^2\left(\frac{1}{2}\Delta + V\right)$. Its lifetime $\tau' := \zeta'(x)$ equals ∞ a.s. and in terms of $X'_r(x)$ we can rewrite $\tau(r)$ by the pathwise integral $\tau(r) = \int_0^r f^2(X'_s(x))\mathrm{d}s$.*

Proof. The last assertion is clear due to $f^2(X_r(x))\mathrm{d}T(r) = \mathrm{d}r$, thus

$$\tau(r) = \int_0^{\tau(r)} f^2(X_s(x))\mathrm{d}T(s) = \int_0^r f^2(X_{\tau(s)}(x))\mathrm{d}s.$$

For the statement concerning L', Itô's formula provides

$$\begin{aligned} g \circ X_{\tau(r)}(x) &\overset{m}{=} \int_0^{\tau(r)} Lg(X_s(x))\mathrm{d}s = \int_0^r Lg(X_s(x))f^2(X_s(x))\mathrm{d}T(s) \\ &= \int_0^r Lg(X_{\tau(s)}(x))f^2(X_{\tau(s)}(x))\mathrm{d}s \end{aligned}$$

for any $g \in C_c^\infty(D)$ which shows that $f^2 L$ is the generator of $X'_r(x)$.
We reproduce the proof of [Th-W], Proposition 2.3, to show that $\tau' \equiv \infty$ a.s.

For this purpose, let $\tau_m := \inf\{s \geq 0 : f(X'_s(x)) \leq \frac{1}{m}\}$, $m \in \mathbb{N}$. Since D is bounded, Lemma 2.22 yields

$$L'f^{-1} = f^2\left(\frac{1}{2}\Delta + V\right)f^{-1} = -\left(\frac{1}{2}\Delta + V\right)f + f^{-1}|\operatorname{grad} f|^2 \leq cf^{-1}$$

for some constant $c > 0$. Now choose m_0 with $f(x) \geq \frac{1}{m_0}$ and observe that for all $r > 0$ and $m \geq m_0$

$$\mathbb{E}f^{-1}(X'_{r\wedge\tau_m}(x)) = f^{-1}(x) + \mathbb{E}\int_0^{r\wedge\tau_m} L'f^{-1}(X'_s(x))\mathrm{d}s \leq f^{-1}(x) + c\int_0^r \mathbb{E}f^{-1}(X'_{s\wedge\tau_m}(x))\mathrm{d}s$$

according to the diffusion property of $X'_r(x)$ (stopping at τ_m assures that the omitted first order term of the Itô formula is a true martingale).
We apply Gronwall's lemma to the continuous function $s \mapsto \mathbb{E}f^{-1}(X'_{s\wedge\tau_m}(x))$ on $[0, r]$ to obtain

$$\mathbb{E}f^{-1}(X'_{r\wedge\tau_m}(x)) \leq f^{-1}(x)e^{cr}, \qquad r > 0, \qquad m \geq m_0.$$

Now $\mathbb{E}f^{-1}(X'_{r\wedge\tau_m}(x)) \geq m\mathbb{P}\{\tau_m < r\}$ implies $\mathbb{P}\{\tau_m < r\} \leq \frac{1}{m}f^{-1}(x)e^{cr}$. Because $\tau' \geq \tau_m$ for any $m \in \mathbb{N}$, we have $\mathbb{P}\{\tau' < r\} = 0$ for all $r > 0$. □

According to these time changes, we choose K_r on the time horizon $[0, t]$ the following way.

Proposition 4.6. *Define a continuous increasing* $\left(\mathscr{F}_r^{X(x)}\right)$*-adapted real-valued process* $(h_0(r))_{r\in\mathbb{R}_+}$ *starting in* 0 *by*

$$h_0(r) := \int_0^r f^{-2}(X_s(x))1_{\{s<\tau(t)\}}\mathrm{d}s \tag{4.13}$$

(which implies $h_0(r) = T(r)$ *for* $r \leq \tau(t)$*). In addition to this, let* $h_1 \in C^1([0, t])$ *denote a real differentiable function with* $h_1(0) = t$ *and* $h_1(t) = 0$*.*
Then by letting $k_r := h_1 \circ h_0(r)$, $0 \leq r \leq t$, *and*

$$K_r := \frac{v}{t}\, k_r = \frac{v}{t}\, h_1 \circ h_0(r) \tag{4.14}$$

we obtain an finite energy process $(K_r)_{r\in[0,t]}$ *that satisfies the assumptions of Theorem 2.27.*
More exactly, we have

$$\left(\int_0^t |\dot{K}_s|^2\mathrm{d}s\right)^{\frac{1}{2}} \in L^p(\mathbb{P}) \quad \textit{for all } p \geq 1. \tag{4.15}$$

Proof. All the claimed facts about h_0 are evident, as well as $k_0 = h_1(0) = t$ and $k_r = 0$ for $r \geq \tau(t)$ since $h_0(\tau(t)) = t$. Hence $K_0 = v$ and $K_r = 1_{[0,\tau(t)]}(r)K_r$ (and $[0, \tau(t)] \subset [0, \tau\wedge t]$). Obviously, it is sufficient to prove (4.15) for $p = 2m$, $m \in \mathbb{N}$, i.e.

$$\mathbb{E}\left(\int_0^t |\dot{K}_s|^2\mathrm{d}s\right)^m < \infty.$$

Due to Jensen's inequality on $[0,t]$, we only need to verify that $\int_0^t |\dot K_s|^{2m}\mathrm{d}s$ has finite expectation.
For this purpose let $c := \sup_{[0,t]} |\dot h_1|^m$ which in combination with $\dot k_r = \dot h_1(h_0(r))\dot h_0(r)$ yields

$$\begin{aligned}\int_0^{\tau(t)} |\dot K_s|^{2m}\mathrm{d}s &\le \frac{c|v|^{2m}}{t^{2m}}\int_0^{\tau(t)} |\dot h_0(s)|^{2m}\mathrm{d}s\\ &\le \frac{c|v|^{2m}}{t^{2m}}\int_0^{\tau(t)} f^{-4m}(X_s(x))\mathrm{d}s\\ &= \frac{c|v|^{2m}}{t^{2m}}\int_0^{\tau(t)} f^{-4m+2}(X_s(x))\mathrm{d}T(s)\\ &= \frac{c|v|^{2m}}{t^{2m}}\int_0^{t} f^{-4m+2}(X'_s(x))\mathrm{d}s\end{aligned}$$

by switching time scale from s to $\tau(s)$ in the last step.
To estimate the expectation of the last integral, we give an argument based on the diffusion property of $X'_r(x)$ and the Gronwall lemma. This method recurs throughout the rest of the chapter with several modifications; we will refer to the proof here for details.
Lemma 2.22 provides

$$\begin{aligned}L'f^{-4m+2} &= f^2\left(\frac{1}{2}\Delta + V\right)f^{-4m+2}\\ &= f^2\left(\frac{(4m-2)(4m-1)}{2}f^{-4m}|\operatorname{grad} f|^2 - (4m-2)f^{-4m+1}Lf\right)\\ &\le c_0(f)f^{-4m+2},\end{aligned}$$

where $c_0(f) := \sup_D\left(\frac{1}{2}(4m-2)(4m-1)|\operatorname{grad} f|^2 - (4m-2)fLf\right)_+ < \infty$.
By introducing stopping times $\sigma_l := \inf\{s \ge 0 : f^{-4m+2}(X'_s(x)) \ge l\}$ for $l \in \mathbb{N}$ we assure that the process $f^{-4m+2}(X'_{s\wedge\sigma_l}(x))$ is bounded which implies by the diffusion property (i.e. application of Itô's formula)

$$\begin{aligned}\mathbb{E}f^{-4m+2}(X'_{s\wedge\sigma_l}(x)) &= f^{-4m+2}(x) + \mathbb{E}\int_0^{s\wedge\sigma_l} L'f^{-4m+2}(X'_{u\wedge\sigma_l}(x))\mathrm{d}u\\ &\le f^{-4m+2}(x) + c_0(f)\int_0^s \mathbb{E}f^{-4m+2}(X'_{u\wedge\sigma_l}(x))\mathrm{d}u.\end{aligned}$$

Hence using Gronwall we find

$$\mathbb{E}f^{-4m+2}(X'_{s\wedge\sigma_l}(x)) \le f^{-4m+2}(x)e^{c_0(f)s}, \quad \text{for all } l \text{ with } f^{-4m+2}(x) < l.$$

By Fatou's lemma we can overcome stopping at σ_l:

$$\begin{aligned}\mathbb{E}f^{-4m+2}(X'_s(x)) &= \mathbb{E}\lim_{l\to\infty} f^{-4m+2}(X'_{s\wedge\sigma_l}(x))\\ &\le \liminf_{l\to\infty}\mathbb{E}f^{-4m+2}(X'_{s\wedge\sigma_l}(x)) \le f^{-4m+2}(x)e^{c_0(f)s}.\end{aligned}$$

So finally we arrive at

$$\mathbb{E}\int_0^{\tau(t)} |\dot K_s|^{2m}\mathrm{d}s \le \frac{c|v|^{2m}}{t^{2m}}\int_0^t \mathbb{E}f^{-4m+2}(X'_s(x))\mathrm{d}s \le \frac{c|v|^{2m}}{t^{2m}}\int_0^t e^{c_0(f)s}\mathrm{d}s < \infty.$$

□

To estimate the stochastic integral in (2.37), we need an upper bound for the operator norm of the damped transport W_r along the paths of $X_r(x)$, $r \le \tau$.

Lemma 4.7. *For k_V given as in (4.6), we have*

$$\|W_r\| \le e^{-k_V r/2}, \qquad r \in [0,\tau]. \tag{4.16}$$

Proof. The defining pathwise equation for W_r

$$\frac{\nabla}{\partial r} W_r = -\frac{1}{2}\operatorname{Ric}(W_r,\,\cdot\,)^\sharp + \nabla V(W_r), \quad W_0 = \mathrm{id}_{T_xM}$$

implies

$$\begin{aligned}\frac{\partial}{\partial r}\|W_r\|^2 &= \left\langle 2W_r, \frac{\nabla}{\partial r}W_r\right\rangle = \langle W_r, -\operatorname{Ric}(W_r,\,\cdot\,)^\sharp + 2\nabla V(W_r)\rangle \\ &= -\operatorname{Ric}(W_r,W_r) + 2\langle \nabla V(W_r), W_r\rangle \le -k_V\|W_r\|^2.\end{aligned}$$

Since $\|W_0\| = 1$ the Gronwall lemma instantly provides $\|W_r\|^2 \le e^{-k_V r}$. □

To derive Hessian estimates in the following section, we will need a similar upper bound for $\|W_r^{-1}\|$.

We finish the proof of Theorem 4.1 by the following statement.

Proposition 4.8. *Consider $v \in T_xM$, $|v| = 1$,*

$$h_1(r) := t - \frac{tc(f)}{1-e^{-c(f)t}}\int_0^r e^{-c(f)s}\mathrm{d}s, \tag{4.17}$$

which obviously is C^1 with $h_1(0) = t$ and $h_1(t) = 0$, and let

$$K_r = \frac{v}{t}k_r = \frac{v}{t}h_1 \circ h_0(r)$$

according to (4.14) and (4.13).
Then for $c(f)$ given by (4.5)

$$\left\|\int_0^{\tau\wedge t}\langle W_r\dot K_r, //_{0,r}\mathrm{d}B_r\rangle\right\|_{L^2(\mathbb{P})} \le \sqrt{\frac{c(f)}{f^2(x)(1-e^{-c(f)t})}}. \tag{4.18}$$

Proof. We rewrite the stochastic integral as $\int_0^{\tau\wedge t}\langle I_r, \mathrm{d}B_r\rangle_{T_xM}$ with $I_r := //_{0,r}^{-1}W_r\dot K_r$. By the foregoing lemma and an estimate on $|\dot K_r|$ carried out below, it is clear that I_r on $[0,\tau\wedge t]$ is square integrable with respect to the Doléans measure $n\mathrm{d}r \otimes \mathbb{P}$ of B_r.
For this reason $\int\langle I_r, \mathrm{d}B_r\rangle$ on $[0,\tau\wedge t]$ is an L^2-bounded martingale and the expectation of its square equals the expectation of its quadratic variation process. So actually we have

$$\begin{aligned}\mathbb{E}\left[\int_0^{\tau\wedge t}\langle I_r, \mathrm{d}B_r\rangle\right]^2 &= \mathbb{E}\left[\int_0^{\tau\wedge t}|I_r|^2\mathrm{d}r\right] \le \mathbb{E}\left[\int_0^{\tau\wedge t}\|W_r\|^2\,|\dot K_r|^2\mathrm{d}r\right] \\ &\le \frac{1}{t^2}\mathbb{E}\left[\int_0^{\tau(t)} e^{-k_V r}\dot h_1^2(h_0(r))f^{-4}(X_r(x))\mathrm{d}r\right]\end{aligned}$$

$$= \frac{1}{t^2}\mathbb{E}\left[\int_0^{\tau(t)} e^{-k_V r}\dot{h}_1^2(h_0(r))f^{-2}(X_r(x))\mathrm{d}T(r)\right]$$
$$= \frac{1}{t^2}\int_0^t \dot{h}_1^2(r)\mathbb{E}\left[f^{-2}(X_r'(x))e^{-k_V\tau(r)}\right]\mathrm{d}r$$

where in the last step we have replaced r with $\tau(r)$ (recall $h_0(\tau(r)) = T(\tau(r)) = r$). As above, the expectation on the right hand side can be estimated by a Gronwall argument. Itô's formula applied to the time-changed diffusion provides

$$\begin{aligned}\mathrm{d}\left(f^{-2}(X_r'(x))e^{-k_V\tau(r)}\right) &\overset{m}{=} f^{-2}(X_r'(x))\mathrm{d}e^{-k_V\tau(r)} + e^{-k_V\tau(r)}L'f^{-2}(X_r'(x))\mathrm{d}r\\ &= -k_V f^{-2}(X_r'(x))e^{-k_V\tau(r)}f^2(X_r'(x))\mathrm{d}r\\ &\quad + f^2(X_r'(x))\left(3f^{-4}|\operatorname{grad} f|^2 - 2f^{-3}Lf\right)(X_r'(x))e^{-k_V\tau(r)}\mathrm{d}r\\ &\le c(f)f^{-2}(X_r'(x))e^{-k_V\tau(r)}\mathrm{d}r,\end{aligned}$$

where the constant is given by $c(f) = \sup_D\left(-f^2k_V + 3|\operatorname{grad} f|^2 - 2fLf\right)_+$, cf. (4.5), which of course is finite due to the boundedness of D and $f \in C^2(\bar{D})$. For the rest of the proof we assume $c(f) > 0$, otherwise we use $c(f) + \varepsilon > 0$ and let $\varepsilon \searrow 0$ afterwards.
The local martingale that both sides of the integrated version of the first equality differ by is a L^2-bounded martingale on every stochastic interval $[0, r \wedge \sigma_l]$, σ_l as in the proof of Proposition 4.6 (with $m = 1$), so we can integrate the derived inequality from 0 to $r \wedge \sigma_l$ and take the expectation on both sides. Then the Gronwall lemma yields for $r \le t$ and l big enough that

$$\mathbb{E}\left[f^{-2}(X_{r\wedge\sigma_l}'(x))e^{-k_V\tau(r\wedge\sigma_l)}\right] \le f^{-2}(x)e^{c(f)r}.$$

By the same Fatou argument as in the proof of Proposition 4.6 we can omit stopping at σ_l on the left hand side.
Taking into account the choice of $h_1(r)$ in (4.17) we finish the proof by

$$\begin{aligned}\mathbb{E}\left[\int_0^{\tau\wedge t}\langle I_r, \mathrm{d}B_r\rangle\right]^2 &\le \frac{c^2(f)}{(1-e^{-c(f)t})^2}\int_0^t \mathrm{e}^{-2c(f)r}f^{-2}(x)e^{c(f)r}\mathrm{d}r\\ &= \frac{c(f)}{f^2(x)(1-e^{-c(f)t})}.\end{aligned}$$

□

The remaining task for this treatment of gradient estimates is to give an upper bound based on an explicit choice of f and D. The following theorem can be found in [Th-W] as Corollary 5.1.

Theorem 4.9. *Let still $D \subset M$ denote an open regular domain, $n = \dim(M)$, and $L = \frac{1}{2}\Delta + V$. For brevity, we write $r(y) := \operatorname{dist}(y, \partial D)$, $y \in D$.*
Let k_V and k_0 by given by (4.6) (the latter one to the case $V = 0$) and for $r > 0$ let

$$C(r) := \frac{1}{2}\sqrt{\frac{\pi^2(n+3)}{r^2} + \frac{2\pi\left(\sqrt{(-k_0)_+(n-1)} + 2\sup_D|V|\right)}{r} + 4(-k_V)_+}. \tag{4.19}$$

Then for $F : [0,t] \times M \to \mathbb{R}$ space-time-harmonic on $[0,t] \times \bar{D}$, we have the gradient estimate

$$|\operatorname{grad}_x F_0| \le \frac{C(r(x))}{\sqrt{1-e^{-C(r(x))^2 t}}} \left(\|F_t\|_D \vee \sup_{[0,t]\times\partial D} |F| \right), \qquad x \in D. \tag{4.20}$$

In the special case of $F_s = u$ being L-harmonic, the estimate simplifies to

$$|\operatorname{grad}_x u| \le C(r(x)) \|u\|_D, \qquad x \in D. \tag{4.21}$$

Proof. For fixed $x \in D$ write $\delta(x) := r(x) \wedge \operatorname{dist}(x, \operatorname{cut}(x)) > 0$. We want to apply Theorem 4.1 to the open ball $B := B_{\delta(x)}(x)$ instead of D in the particular case of f defined by

$$f(p) := \cos\left(\frac{\pi \operatorname{dist}(x,p)}{2\delta(x)}\right) \equiv \bar{f}(\operatorname{dist}(x,p)), \qquad p \in \bar{B},$$

$\bar{f}(r) := \cos(\pi r/(2\delta(x)))$, $r \ge 0$. The restriction to a ball inside the cut-locus of x implies $f \in C^2(B)$, f continuous and vanishing on ∂B, and $0 < f|_B \le 1$ with $f(x) = 1$. Moreover, on B we have $|\operatorname{grad} f| \le \frac{\pi}{2\delta(x)}$.
Without loss of generality, we may assume for the rest of the proof that $(-k_0)_+ > 0$. Otherwise, take $\varepsilon > 0$ instead of $(-k_0)_+$ and let $\varepsilon \searrow 0$ afterwards.
Comparison of the Laplacian of the distance from x on B with that on a model space $\mathbb{M}$ of constant radial curvature $k_{\mathbb{M}}(r) = -\frac{(-k_0)_+}{n-1}$ yields

$$\Delta \operatorname{dist}(x,\,\cdot\,)(p) \le \sqrt{(n-1)(-k_0)_+} \coth\left(\sqrt{\frac{(-k_0)_+}{n-1}} \operatorname{dist}(x,p)\right).$$

We are going to explain this in a short remark below.
Now the composition rule tells us that

$$\begin{aligned}
-\Delta f(p) &= -\Delta(\bar{f} \circ \operatorname{dist}(x,\,\cdot\,))(p) \\
&= -\bar{f}'(\operatorname{dist}(x,p)) \Delta \operatorname{dist}(x,\,\cdot\,)(p) - \bar{f}''(\operatorname{dist}(x,p)) |\operatorname{grad} \operatorname{dist}(x,\,\cdot\,)|^2(p) \\
&\le \frac{\pi\sqrt{(n-1)(-k_0)_+}}{2\delta(x)} \sin\left(\frac{\pi \operatorname{dist}(x,p)}{2\delta(x)}\right) \coth\left(\sqrt{\frac{(-k_0)_+}{n-1}} \operatorname{dist}(x,p)\right) \\
&\qquad + \frac{\pi^2}{4\delta(x)^2} \cos\left(\frac{\pi \operatorname{dist}(x,p)}{2\delta(x)}\right)
\end{aligned}$$

since $|\operatorname{grad} \operatorname{dist}(x,\,\cdot\,)| \equiv 1$. Well known from calculus are $\sin r \le 1 \wedge r$ and $\coth r \le 1 + \frac{1}{r}$, $r > 0$. Hence

$$\sin\left(\frac{\pi \operatorname{dist}(x,p)}{2\delta(x)}\right) \coth\left(\sqrt{\frac{(-k_0)_+}{n-1}} \operatorname{dist}(x,p)\right) \le 1 + \frac{\pi\sqrt{n-1}}{2\delta(x)\sqrt{(-k_0)_+}}.$$

So we can put the estimate (derived by $\cos \le 1$ in the right hand side of the upper bound of $-\Delta f$)

$$-\Delta f \le \frac{\pi\sqrt{(-k_0)_+(n-1)}}{2\delta(x)} + \frac{\pi^2 n}{4\delta(x)^2}$$

into the definition of $c(f)$ and find

$$c(f) \leq (-k_V)_+ + \frac{(3+n)\pi^2}{4\delta(x)^2} + \frac{\pi\sqrt{(-k_0)_+(n-1)} + 2\sup_D |V|}{2\delta(x)} = C(\delta(x))^2.$$

Herein we have used

$$\begin{aligned} |Vf|(p) = |\langle \operatorname{grad} f, V\rangle|(p) &\leq \left|\frac{\pi}{2\delta(x)} \sin\left(\frac{\pi \operatorname{dist}(x,p)}{2\delta(x)}\right)\right| |\operatorname{grad}\operatorname{dist}(x,\,\cdot\,)(p)| \sup_D |V| \\ &\leq \frac{\pi}{2\delta(x)} \sup_D |V|. \end{aligned}$$

So the proof is finished if we are allowed to replace $\delta(x)$ by $r(x)$, which means that the result does not depend on the cut-locus of x. This is done by an observation of Kendall, cf. [Ke], which can be also found in [H-Th], 7.254:
For any $o \in D = B_{r(x)}$ the real process $\operatorname{dist}(o, X_r(x))$ on $[0,\tau]$ has the property

$$\operatorname{dist}(o, X_r(x)) - \operatorname{dist}(o,x) = \widehat{M}_r + \int_0^r L\operatorname{dist}(o,\,\cdot\,)(X_s(x))\mathrm{d}s - L_r^{(o)}. \tag{4.22}$$

Herein, by convention, $L\operatorname{dist}(o,\,\cdot\,) = 0$ on the subset of M where the distance fails to be differentiable; $\widehat{M}_r$ is a martingale and $L_r^{(o)}$ is an increasing adapted process with $1_{\{X_r(x)\notin \operatorname{cut}(o)\}}(r)\mathrm{d}L_r^{(o)} = 0$. Returning to the proof of Proposition 4.8, this representation leads to

$$\begin{aligned} \mathrm{d}f^{-2}(X_{\tau(r)}(x)) &= \mathrm{d}\left(\bar{f}^{-2} \circ (\operatorname{dist}(x, X_{\tau(r)}(x))\right) \\ &= \mathrm{d}\widetilde{M}_{\tau(r)} + L' f^{-2}(X_{\tau(r)}(x)) - (\bar{f}^{-2})'\mathrm{d}L_{\tau(r)}^{(o)}, \end{aligned}$$

where $\widetilde{M}_{\tau(r)}$ is again a martingale. Since $\bar{f}$ is decreasing on $[0,\delta(x)]$, the first derivative of $\bar{f}^{-2}$ is positive, and thus we have the same estimate

$$\mathrm{d}\left(f^{-2}(X_{\tau(r)}(x))\right) \leq \mathrm{d}\widetilde{M}_{\tau(r)} + L' f^{-2}(X_{\tau(r)}(x))$$

as we used in the proposition. □

Remark 4.10 *(On the comparison result for* $\Delta\operatorname{dist}(x,\,\cdot\,)$ *on* B*).*
To get the upper bound for $\Delta\operatorname{dist}(x,\,\cdot\,)(p)$ as above, we make use of the following version of the *Laplacian comparison theorem*, cf. [H-Th], Satz 7.243:
If $\operatorname{Ric}_p(v,v) \geq (n-1)c$ for a constant $c \in \mathbb{R}$ and for all $p \in B$ and $v \in T_pM$ with $|v| = 1$, then

$$\Delta\operatorname{dist}(x,\,\cdot\,)(p) \leq \Delta\operatorname{dist}_{\mathbb{M}}(0,\,\cdot\,)(\tilde{p})$$

where $\mathbb{M}$ is a rotationally symmetric manifold (a so-called *model*) with centre 0 and constant radial curvature c, $\dim(\mathbb{M}) = \dim(M) = n \geq 2$ and $\tilde{p} \in \mathbb{M}$ with $\operatorname{dist}_{\mathbb{M}}(0,\tilde{p}) = \operatorname{dist}(x,p)$.
In our situation we have $c = -\frac{(-k_0)_+}{n-1}$ with $(-k_0)_+ > 0$ according to the proof above. Hence we may take for our model $\mathbb{M}$ the *n-dimensional hyperbolic space with curvature* c,

given as $\mathbb{M} := \{x \in \mathbb{R}^n : |x| < \frac{1}{\sqrt{|c|}}\}$, where in polar coordinates the Riemannian metric on $\mathbb{M}$ can be written in the form

$$d\rho \otimes d\rho + g^2(\rho) d\vartheta^2$$

with $\rho(\tilde{p}) = \mathrm{dist}_{\mathbb{M}}(0, \tilde{p})$ and $g(\rho) = \frac{1}{\sqrt{|c|}} \sinh\left(\rho\sqrt{|c|}\right)$. As an elementary property of models (cf. [H-Th], Satz 7.244), we can compute $\Delta_{\mathbb{M}} \mathrm{dist}_{\mathbb{M}} = (n-1)\frac{g'}{g} \circ \mathrm{dist}_{\mathbb{M}}$ which yields the desired estimate.

We finish this paragraph by giving another function g that could have been used instead of $f = \cos\left(\frac{\pi\,\mathrm{dist}(x,p)}{2\delta(x)}\right)$ in the proof of Theorem 4.9 which gives a slightly worse constant, but is easier to transfer to the Hessian case below.

Remark 4.11. We retain the notation of Theorem 4.9, but now use the function $g(p) = 1 - \left(\frac{\mathrm{dist}(x,p))}{\delta(x)}\right)^3 = \bar{g} \circ \mathrm{dist}(x,\,\cdot\,)(p)$ on $B_{\delta(x)}(x)$, where $\bar{g}(r) := 1 - \frac{r^3}{\delta(x)^3}$. Then the same computations as in the proof there, in particular the estimate on the Laplacian of the distance function as well as $\coth(r) \leq 1 + \frac{1}{r}$ and $\frac{\mathrm{dist}(x,p)}{\delta(x)} \leq 1$ on $B_{\delta(x)}(x)$, yield

$$-(\Delta g)(p) \leq \frac{3\sqrt{(n-1)(-k_0)_+}}{\delta(x)} + \frac{3(n+1)}{\delta(x)^2}.$$

Consequently, we derive

$$c(g) \leq (-k_V)_+ + \frac{3(n+4)}{\delta(x)^2} + \frac{3\sqrt{(n-1)(-k_0)_+} + 6\sup_D |V|}{\delta(x)}$$

and by this we get the result of the theorem with the slightly increased constant

$$C(r) := \sqrt{\frac{3(n+4)}{r^2} + \frac{3\left(\sqrt{(-k_0)_+(n-1)} + 6\sup_D |V|\right)}{r} + (-k_V)_+}\,.$$

4.2. Local Hessian estimate

We are now going to transfer the gradient estimate of Thalmaier and Wang to the Hessian case by exploiting our formula (3.33). According to our assumptions in the Hessian representation theorem, we only treat the case of $V = 0$, i.e. $X_r(x)$ is a Brownian motion on M starting in x with generator $L = \frac{1}{2}\Delta$.

Since in the intrinsic Hessian theorem there appear two finite energy processes that vary on disjoint stochastic time intervals we will also need two relatively compact neighbourhoods D and $\widetilde{D}$ of x such that $\bar{D} \subset \widetilde{D}$ as well as two C^2-functions f and $\tilde{f}$ the main theorem depends on. We will distinguish between functions, processes and domains related to the first “half” of the time interval and the full time scale by writing $\tilde{}$ on top of the latter ones.

As the Hessian is a symmetric bilinear form on the tangent bundle, for an upper bound on its norm it is sufficient to estimate its values on the diagonal.

Theorem 4.12. *Let $X_r(x)$ denote a Brownian motion with lifetime $\zeta(x)$ on the Riemannian manifold (M,g) of dimension $n \geq 2$.*
Consider two open regular domains D, $\widetilde{D}$ with $x \in D \subseteq \bar{D} \subset \widetilde{D}$. Let σ be the first exit time of $X_r(x)$ from D, and τ the same with respect to $\widetilde{D}$. Moreover, fix $f \in C^2(\bar{D})$ with $0 < f|_D \leq \frac{1}{\sqrt{2}}$ and $f|_{\partial D} = 0$ as well as $\tilde{f} \in C^2(\bar{\widetilde{D}})$ with $0 < \tilde{f}|_{\widetilde{D}} \leq \frac{1}{\sqrt{2}}$, being constant on D ($\tilde{f}(p) = \tilde{f}(x)$ for all $p \in D$) and $\tilde{f}_{\partial\widetilde{D}} = 0$. As in the gradient case we consider an arbitrary $t > 0$ and $F : [0,t] \times M \to \mathbb{R}$, which is space-time-harmonic on $[0,t] \times \bar{\widetilde{D}}$.
In addition to this, we denote curvature bounds on D by

$$C_1 := \sup_{y \in D,\, w \in T_yM,\, |w|=1} |(\nabla \operatorname{Ric}^\sharp + d^*R)(w,w)| \tag{4.23}$$

and

$$C_2 := \sup_{y \in D,\, w,w' \in T_yM,\, |w|=|w'|=1} |R(w,w')w'|. \tag{4.24}$$

Furthermore, let k and K be two real constants satisfying

$$k \leq \inf_{y \in \widetilde{D}, w \in T_yM,\, |w|=1} \operatorname{Ric}(w,w) \leq \sup_{y \in D,\, w \in T_yM,\, |w|=1} \operatorname{Ric}(w,w) \leq K. \tag{4.25}$$

We assume $2k \neq K$ (otherwise we may choose $K + \varepsilon$ instead of K).
Then for any $v \in T_xM$ with $|v| = 1$ we have the inequality

$$\begin{aligned} |\operatorname{Hess}_x F_0| &\equiv \sup_{v \in T_xM,\, |v|=1} |\operatorname{Hess}_x F_0(v,v)| \leq C \left(\|F_t\|_D \vee \sup_{[0,t]\times\partial D} |F| \right), \\ C &:= \frac{\sqrt{c(f)}}{f(x)\sqrt{1-e^{-c(f)t}}} \left(\frac{C_1}{|K-2k|} + \frac{C_2 C_4^{1/4}\sqrt{n}}{\sqrt{|K-2k|}} + \frac{(1 \vee e^{-kt/4})\sqrt{\tilde{c}(\tilde{f})}}{\tilde{f}(x)\sqrt{1-e^{-\tilde{c}(\tilde{f})t}}} \right), \end{aligned} \tag{4.26}$$

hence $C = C(D, \widetilde{D}, f, \tilde{f}, t, n)$.
More exactly, C_4 denotes the upper constant in the Burkholder-Davis-Gundy inequality for $p = 4$, and $c(f)$ and $\tilde{c}(\tilde{f})$ are positive constants depending on f (resp. $\tilde{f}$), its derivatives and k, K. Actually, they may be chosen as

$$c(f) := \sup_D \left(-kf^2 + 5|\operatorname{grad} f|^2 - f\Delta f\right)_+ + (K-2k)_+ \sup_D f^2. \tag{4.27}$$

and

$$\tilde{c}(\tilde{f}) := \sup_{\widetilde{D}} (-k\tilde{f}^2 + 3|\operatorname{grad} \tilde{f}|^2 - \tilde{f}\Delta\tilde{f})_+ \tag{4.28}$$

If $K < 2k$, $c(f)$ and $\tilde{c}(\tilde{f})$ do not depend on K.

Remark 4.13. Observe that all particular situations and extensions concerning the term $\|F_t\|_D \vee \sup_{[0,t]\times\partial D} |F|$ that we discussed in Remark 4.2 still hold true here.
For this reason we stick to this general upper bound and focus on the estimate of the stochastic integrals in the Hessian representation formula.

Proof of Theorem 4.12. Exactly as in the proof of Theorem 4.1 we derive by simple application of the Cauchy-Schwartz inequality to the Hessian representation formula (3.33)

$$\begin{aligned}
|\operatorname{Hess}_x F_0(\,\cdot\,)(v,v)| &\le \left(\|F_t\|_D \vee \sup_{[0,t]\times\partial D} |F|\right) \left\| G_{\sigma\wedge\frac{t}{2}} + \frac{1}{2} H_{\sigma\wedge\frac{t}{2}} + I_{\tau\wedge t} J_{\sigma\wedge\frac{t}{2}} \right\|_{L^2(\mathbb{P})} \\
&\le \left(\|F_t\|_D \vee \sup_{[0,t]\times\partial D} |F|\right) \left[\left(\mathbb{E}[G^2_{\sigma\wedge\frac{t}{2}}]\right)^{\frac{1}{2}} + \frac{1}{2}\left(\mathbb{E}[H^2_{\sigma\wedge\frac{t}{2}}]\right)^{\frac{1}{2}} + \left(\mathbb{E}[I^2_{\tau\wedge t} J^2_{\sigma\wedge\frac{t}{2}}]\right)^{\frac{1}{2}} \right].
\end{aligned}$$

Note that in formula (3.33) we may write $F_{\tau\wedge t}\circ X_{\tau\wedge t}(x)$ in all terms on the right hand side according to the martingale property of $F_s \circ X_s(x)$ and since $\tau\wedge t$ plays the role of τ there.

For brevity, we herein introduced the stochastic integrals G, H, I and J with respect to $\mathrm{d}^\nabla X_r = A(X_r)\mathrm{d}Z_r = //_{0,r}\mathrm{d}B_r$ given by:

$$G_{\sigma\wedge\frac{t}{2}} := -\int_0^{\sigma\wedge\frac{t}{2}} \left\langle W_r \int_0^r W_s^{-1} R\left(\mathrm{d}^\nabla X_s, W_s\, v\right) W_s\, \dot{L}_r, //_{0,r}\mathrm{d}B_r \right\rangle, \tag{4.29}$$

$$H_{\sigma\wedge\frac{t}{2}} := \int_0^{\sigma\wedge\frac{t}{2}} \left\langle W_r \int_0^r W_s^{-1} \left(\nabla \operatorname{Ric}^\sharp + d^* R\right) \left(W_s\, v, W_s\, \dot{L}_r\right) \mathrm{d}s, //_{0,r}\mathrm{d}B_r \right\rangle, \tag{4.30}$$

$$I_{\tau\wedge t} := \int_{\sigma\wedge\frac{t}{2}}^{\tau\wedge t} \left\langle W_r \dot{K}_r, //_{0,r}\mathrm{d}B_r \right\rangle, \tag{4.31}$$

$$J_{\sigma\wedge\frac{t}{2}} := \int_0^{\sigma\wedge\frac{t}{2}} \left\langle W_r \dot{L}_r, //_{0,r}\mathrm{d}B_r \right\rangle. \tag{4.32}$$

The finite energy processes K and L should satisfy the assumptions of Theorem 3.12, in particular $K_0 = L_0 = v$, $K_{\tau\wedge t} = L_{\tau\wedge t} = 0$. Moreover, $\dot{K}_r = 1_{[\sigma\wedge\frac{t}{2},\tau\wedge t]}(r)\dot{K}_r$ and $\dot{L}_r = 1_{[0,\sigma\wedge\frac{t}{2}]}(r)\dot{L}_r$. A construction of both processes depending on D, $\widetilde{D}$, f and $\tilde{f}$ will be given in the proposition below.

For this reason the proof is finished by giving estimates on the expectations of $G^2_{\sigma\wedge\frac{t}{2}}$, $H^2_{\sigma\wedge\frac{t}{2}}$ and $I^2_{\tau\wedge t}J^2_{\sigma\wedge\frac{t}{2}}$ based on the constants of the theorem.

This is done in the Lemmas 4.18, 4.19 and 4.20, respectively. Every single estimate on the expectations related to H, G and IJ (in this order) corresponds to constants

$$c_1(f) := \begin{cases} \sup_D \left((K-3k)f^2 + 3|\operatorname{grad} f|^2 - f\Delta f\right)_+, & K > 2k, \\ \sup_D \left(-kf^2 + 3|\operatorname{grad} f|^2 - f\Delta f\right)_+, & K < 2k, \end{cases} \tag{4.33}$$

$$c_2(f) := \begin{cases} \sup_D \left(-kf^2 + 5|\operatorname{grad} f|^2 - f\Delta f\right)_+ + (K-2k)\sup_D f^2, & K > 2k, \\ \sup_D \left(-kf^2 + 5|\operatorname{grad} f|^2 - f\Delta f\right)_+, & K < 2k, \end{cases} \tag{4.34}$$

and

$$c_3(f) := \sup_D(-kf^2 + 3|\operatorname{grad} f|^2 - f\Delta f)_+. \tag{4.35}$$

However, since we will have to find one constant $c(f)$ which appears in the construction of L_s (analogously to the gradient case), we pass over to their maximum

$$c(f) := c_1(f) \vee c_2(f) \vee c_3(f).$$

Now it is easily seen that in both cases of the relation between K and $2k$ the constant $c_2(f)$ dominates the other ones and this explains the statement of (4.27). □

Definition 4.14. According to the assumptions on D, $\widetilde{D}$, f and $\tilde{f}$ of the theorem, we define time changes

$$T(r) := \int_0^r f^{-2}(X_s(x))\mathrm{d}s, \qquad 0 \le r \le \sigma, \tag{4.36}$$

$$\tau(r) := \inf\{s \ge 0 : T(s) \ge r\}, \qquad r \in \overline{\mathbb{R}}_+, \tag{4.37}$$

$$\widetilde{T}(r) := \int_{(\sigma\wedge\frac{t}{2})\wedge r}^r \tilde{f}^{-2}(X_s(x))\mathrm{d}s, \qquad 0 \le r \le \tau, \tag{4.38}$$

$$\widetilde{\tau}(r) := \inf\{s \ge 0 : \widetilde{T}(s) \ge r\}, \qquad r \in \overline{\mathbb{R}}_+. \tag{4.39}$$

We also introduce the time-changed diffusion processes

$$X'_r(x) := X_{\tau(r)}(x), \quad r \in \mathbb{R}_+, \qquad \text{and} \qquad \widetilde{X}_r(x) := \begin{cases} X_{\sigma\wedge\frac{t}{2}}(x), & r = 0, \\ X_{\widetilde{\tau}(r)}(x), & r > 0. \end{cases} \tag{4.40}$$

Remark 4.15. It is clear that $T(r)$ and $\tau(r)$ are adapted, continuous and inverse to each other, i.e. $\tau(T(r)) = r$ for $r \in [0,\sigma]$ and $T(\tau(r)) = r$, $r \in \mathbb{R}_+$. The same argument as in Lemma 4.5 shows that $\tau(t) < \sigma$ and thus, by the additional assumption that $f \le \frac{1}{\sqrt{2}}$ we are given that $T(s) \ge 2s$ for all $0 < s \le \sigma$ and consequently $\tau(t) \le \sigma \wedge \frac{t}{2}$.
A similar statement holds for $\widetilde{T}(r)$ and $\widetilde{\tau}(r)$ with slight modifications. First, $\widetilde{T}|_{[0,\sigma\wedge\frac{t}{2}]} \equiv 0$ and consequently $\widetilde{\tau}(0) = 0 \ne \sigma \wedge \frac{t}{2} = \widetilde{\tau}(0+)$. On $]0,\infty[$ however, $\widetilde{\tau}(\,\cdot\,)$ is continuous. According to this interval where $\widetilde{T}(\,\cdot\,)$ is constant, the times at which both processes are inverse to each other underly the following restrictions: $\tilde{\tau}(\tilde{T}(r)) = r$ for all $\sigma \wedge \frac{t}{2} \le r \le \tau$ and $\widetilde{T}(\widetilde{\tau}(r)) = r$ for arbitrary r. We have $\widetilde{\tau}(t) < \tau$ and, by noting that $\widetilde{T}(\tau \wedge t) \ge \int_{t/2}^t 2\,\mathrm{d}s \wedge \widetilde{T}(\tau) = t$, we conclude $\widetilde{\tau}(t) \le \tau \wedge t$.
The results of Lemma 4.5 of course also hold in this case. In fact, $X'_r(x)$ is a $\big(\mathscr{F}^{X(x)}_{\tau(r)}\big)$-adapted D-valued diffusion with generator $L' := f^2 L = f^2 \frac{1}{2}\Delta$ and (a.s.) infinite lifetime. We also can rewrite $\tau(r)$ in terms of $X'_s(x)$ as $\tau(r) = \int_0^r f^2(X'_s(x))\mathrm{d}s$.
Analogously, $\widetilde{X}_r(x)$ is adapted to $\big(\mathscr{F}^{X(x)}_{(\sigma\wedge\frac{t}{2})\vee\widetilde{\tau}(r)}\big)$, takes values in $\widetilde{D}$, has (a.s.) infinite lifetime and corresponds to the generator $\widetilde{L} := \tilde{f}^2 L = \tilde{f}^2\frac{1}{2}\Delta$.
Note that the definition of $\widetilde{X}_r(x)$ as $X_{\widetilde{\tau}(r)}(x)$ for all r would have caused a jump at $r = 0$, but since we will only make use of functions of $\tilde{f} \circ X_{\widetilde{\tau}(r)}(x)$, $X_{\sigma\wedge\frac{t}{2}}(x) \in \bar{D}$ and $\tilde{f}$ is assumed to be constant on $\bar{D}$, this would not have had any effect on the proof.

Proposition 4.16. *Define two continuous increasing and* $\big(\mathscr{F}^{X(x)}_r\big)$*-adapted real processes* $h_0(r)$ *and* $\tilde{h}_0(r)$ *by*

$$h_0(r) := \int_0^r f^{-2}(X_s(x))1_{\{s<\tau(t)\}}\mathrm{d}s \tag{4.41}$$

and

$$\tilde{h}_0(r) := \int_{(\sigma\wedge\frac{t}{2})\wedge r}^r \tilde{f}^{-2}(X_s(x))1_{\{s<\widetilde{\tau}(t)\}}\mathrm{d}s. \tag{4.42}$$

According to the notions of Theorem 4.12, we obtain two functions

$$h_1(r) := t - \frac{tc(f)}{1-e^{-c(f)t}} \int_0^r e^{-c(f)s} \mathrm{d}s \tag{4.43}$$

and

$$\tilde{h}_1(r) := t - \frac{t\tilde{c}(\tilde{f})}{1-e^{-\tilde{c}(\tilde{f})t}} \int_0^r e^{-\tilde{c}(\tilde{f})s} \mathrm{d}s \tag{4.44}$$

in $C^1([0,t])$.
By means of these definitions let $\ell_r := h_1 \circ h_0(r)$ *and* $k_r := \tilde{h}_1 \circ \tilde{h}_0(r)$, $0 \le r \le t$, *as well as*

$$K_r := \frac{v}{t} k_r = \frac{v}{t} \tilde{h}_1 \circ \tilde{h}_0(r) \qquad \text{and} \qquad L_r = \frac{v}{t} \ell_r = \frac{v}{t} h_1 \circ h_0(r). \tag{4.45}$$

This yields two adapted bounded processes with absolutely continuous paths taking values in T_xM. *We have* $K_0 = L_0 = v$, $K_t = L_t = 0$, $\dot{K}_r = 1_{[\sigma\wedge\frac{t}{2},\tau\wedge t]}(r)\dot{K}_r$ *and* $\dot{L}_r = 1_{[0,\sigma\wedge\frac{t}{2}]}(r)\dot{L}_r$. *The necessary integrability condition* $\int_0^t |\dot{K}_r|^2 \mathrm{d}r \in L^{1+\alpha}(\mathbb{P})$ *(and the same for* L_r*) is e.g. satisfied for* $\alpha = 1$.

Proof. All assertions are evident by the foregoing remark and the definitions of the proposition except for the last integrability statement. This can be verified by reproducing the proof of Proposition 4.6 for the exponent $p = 4$ (in fact, we could show integrability with respect to any $\alpha > 0$). All the arguments carry over verbatim to the situation here, except for replacing f by $\tilde{f}$, L' by $\widetilde{L}$, $\tau(t)$ by $\widetilde{\tau}(t)$, etc. in the case of K_s. □

Since the Hessian Bismut formula contains the inverse damped transport W_r^{-1} as well, we also need an estimate on its norm, which now – in contrast to the case of W_r itself – is related to an upper bound on the Ricci curvature on $\widetilde{D}$.

Lemma 4.17. *According to the assumptions*

$$k \le \inf_{y\in\widetilde{D}, w\in T_yM,\, |w|=1} \mathrm{Ric}(w,w) \le \sup_{y\in D,\, w\in T_yM,\, |w|=1} \mathrm{Ric}(w,w) \le K.$$

in Theorem 4.12, we have the following estimates on the damped and inverse damped transports along the paths of $X_r(x)$:

$$\|W_r\| \le e^{-kr/2}, \qquad 0 \le r \le \tau, \tag{4.46}$$

and

$$\|W_r^{-1}\| \le e^{Kr/2}, \qquad 0 \le r \le \sigma. \tag{4.47}$$

Proof. Equation (4.46) is just the statement of Lemma 4.7 in the case of $V = 0$. As in the proof there one also has

$$\frac{\partial}{\partial r}|W_r w|^2 = -\mathrm{Ric}(W_r w, W_r w) \ge -K|W_r w|^2, \qquad |W_0 w|^2 = |w|^2,$$

for all $w \in W$ and all $0 < r < \sigma$. This implies $|W_r w|^2 \ge |w|^2 e^{-Kr}$ for all $w \in T_xM$ and thus (4.47) holds. □

Using these results, we verify the estimates on the L^2-norms of the stochastic integrals in the Hessian representation formula. We start out by the slightly easier case of $H_{\sigma\wedge\frac{t}{2}}$.

Lemma 4.18. *Under the assumptions of Theorem 4.12 and with L_r being given by the explicit construction in Proposition 4.16, for $H_{\sigma\wedge\frac{t}{2}}$ from (4.30) holds*

$$\mathbb{E}[H^2_{\sigma\wedge\frac{t}{2}}] \leq \frac{4C_1^2 c(f)}{(K-2k)^2 f^2(x)(1-e^{-c(f)t})}\,. \tag{4.48}$$

Proof. We adopt the convention in this and the following two proofs that we assume all constants $c(f)$, $c_i(f)$ $(i = 1, 2, 3)$ and $\tilde{c}(\tilde{f})$ to be greater than 0. If one of them vanishes the arguments can be carried out for a slightly increased constant and one has to take its limit to 0 afterwards.

Our assumptions provide that the process H is a $L^2(\mathbb{P})$-bounded martingale on $[0, \sigma\wedge\frac{t}{2}]$, the L^2-norm of its integrand with respect to the Doléans measure belonging to B_r is implicitly estimated below. For this reason the expecation of $H^2_{\sigma\wedge\frac{t}{2}}$ equals the expectation of its quadratic variation process at time $\sigma\wedge\frac{t}{2}$.
From $|v| = 1$, $\tau(t) \leq \sigma\wedge\frac{t}{2}$ and $\dot{\ell}_r = 0$ for $r \geq \tau(t)$ we conclude

$$\begin{aligned}
\mathbb{E}[H^2_{\sigma\wedge\frac{t}{2}}] &= \mathbb{E}\left[\int_0^{\sigma\wedge\frac{t}{2}} \left|W_r \frac{1}{t}\dot{\ell}_r \int_0^r W_s^{-1}(\nabla \operatorname{Ric}^\sharp + d^*R)(W_s v, W_s v)\mathrm{d}s\right|^2 \mathrm{d}r\right] \\
&\leq \mathbb{E}\left[\int_0^{\sigma\wedge\frac{t}{2}} \|W_r\|^2 \frac{1}{t^2}\dot{\ell}_r^2 \left(\int_0^r \|W_s^{-1}\|\; C_1\; \|W_s\|^2 \mathrm{d}s\right)^2 \mathrm{d}r\right] \\
&\leq \frac{C_1^2}{t^2}\mathbb{E}\left[\int_0^{\tau(t)} (e^{-kr/2})^2 (\dot{h}_1\circ h_0(r))^2 f^{-4}(X_r)(x) \left(\int_0^r e^{Ks/2}(e^{-ks/2})^2\,\mathrm{d}s\right)^2 \mathrm{d}r\right] \\
&= \frac{C_1^2}{t^2}\mathbb{E}\left[\int_0^{\tau(t)} e^{-kr} i^2(r) \dot{h}_1^2(h_0(r)) f^{-2}(X_r(x))\mathrm{d}T(r)\right] \\
&= \frac{C_1^2}{t^2}\mathbb{E}\left[\int_0^{t} e^{-k\tau(r)} i^2(\tau(r)) \dot{h}_1^2(r) f^{-2}(X'_r(x))\,\mathrm{d}r\right] \\
&= \frac{C_1^2}{t^2}\int_0^t \dot{h}_1^2(r)\mathbb{E}\left[e^{-k\tau(r)} i^2(\tau(r)) f^{-2}(X'_r(x))\right]\mathrm{d}r,
\end{aligned}$$

where $i(r) := \int_0^r e^{(K-2k)s/2}\mathrm{d}s = \frac{2}{K-2k}(e^{(K-2k)r/2}-1)$. Because of $k \leq K$ the case $K < 2k$ can only occur if $k > 0$. According to this we have the inequalities

$$e^{-k\tau(r)} i^2(\tau(r)) \leq \begin{cases} \frac{4}{(K-2k)^2} e^{(K-3k)\tau(r)}, & K > 2k, \\ \frac{4}{(2k-K)^2} e^{-k\tau(r)}, & 0 \leq k \leq K < 2k. \end{cases}$$

As in the proof of Proposition 4.8 the Itô formula yields for $K > 2k$

$$\begin{aligned}
\mathrm{d}\left[e^{(K-3k)\tau(r)} f^{-2}(X'_r(x))\right] &\overset{m}{=} (K-3k)e^{(K-3k)\tau(r)}\mathrm{d}r + e^{(K-3k)\tau(r)} L' f^{-2}(X'_r(x))\mathrm{d}r \\
&\leq c_1(f) e^{(K-3k)\tau(r)} f^{-2}(X'_r(x)),
\end{aligned}$$

$c_1(f)$ given in (4.33). The same argument in the case of $K < 2k$ shows

$$\begin{aligned} \mathrm{d}\left[e^{-k\tau(r)} f^{-2}(X_r'(x))\right] &\overset{m}{=} -ke^{-k\tau(r)}\mathrm{d}r + e^{-k\tau(r)} L' f^{-2}(X_r'(x))\mathrm{d}r \\ &\leq c_1(f) e^{-k\tau(r)} f^{-2}(X_r'(x)). \end{aligned}$$

By exactly the same reasoning as in the proposition for the gradient case cited above, the difference in the integrated versions of the first equality is a true martingale (with initial value 0) when stopped at the first exit of $X_r(x)$ from $\{f^2 > \frac{1}{l}\}$ such that its expectation equals 0 as well. Thus, the Gronwall lemma and passing over to $l \to \infty$ according to Fatou (cf. proof of 4.6) yields

$$\begin{aligned} \mathbb{E}\left[e^{(K-3k)\tau(r)} f^{-2}(X_r'(x))\right] &\leq f^{-2}(x) e^{c_1(f)r} \leq f^{-2}(x) e^{c(f)r}, && K > 2k, \\ \mathbb{E}\left[e^{-k\tau(r)} f^{-2}(X_r'(x))\right] &\leq f^{-2}(x) e^{c(f)r}, && K < 2k. \end{aligned}$$

The necessity of the transition from $c_1(f)$ to $c(f)$ is described in the proof of Theorem 4.12 above.
So finally we are given

$$\begin{aligned} \mathbb{E}[H^2_{\sigma\wedge\frac{t}{2}}] &\leq \frac{4C_1^2}{(K-2k)^2 t^2} \int_0^t \dot{h}_1^2(r) f^{-2}(x) e^{c(f)r} \mathrm{d}r \\ &= \frac{4C_1^2 c(f)}{(K-2k)^2 f^2(x)(1-e^{-c(f)t})}. \end{aligned}$$

□

Lemma 4.19. *Still by the same assumptions as in the foregoing lemma the result for the expectation of* $G_{\sigma\wedge\frac{t}{2}}$ *from (4.29) reads*

$$\mathbb{E}[G^2_{\sigma\wedge\frac{t}{2}}] \leq \frac{n\sqrt{C_4}C_2^2 c(f)}{|K-2k| f^2(x)(1-e^{-c(f)t})}\ . \tag{4.49}$$

Proof. Just as in Lemma 4.18 we know that G is an $L^2(\mathbb{P})$-bounded martingale on $[0, \sigma\wedge\frac{t}{2}]$ such that the expectation of its square is given by the expectation of the related quadratic variation process.
Since we only treat the case of a Brownian motion, we have $\mathrm{d}^\nabla X_s = //_{0,s}\mathrm{d}B_s$ and so we start out by

$$\begin{aligned} \mathbb{E}[G^2_{\sigma\wedge\frac{t}{2}}] &= \mathbb{E}\left[\int_0^{\sigma\wedge\frac{t}{2}} \left|\frac{1}{t}\dot{\ell}_r W_r \int_0^r \left(W_s^{-1} R(\,\cdot\,, W_s v) W_s v\right) //_{0,s} dB_s\right|^2 dr\right] \\ &\leq \frac{1}{t^2}\mathbb{E}\left[\int_0^{\tau(t)} \|W_r\|^2 (\dot{h}_1 \circ h_0(r))^2 f^{-4}(X_r)(x) \left|\int_0^r R_s \mathrm{d}B_s\right|^2 dr\right] \\ &\leq \frac{1}{t^2}\mathbb{E}\left[\int_0^{\tau(t)} e^{-kr} (\dot{h}_1 \circ h_0(r))^2 f^{-2}(X_r)(x) \left|\int_0^r R_s \mathrm{d}B_s\right|^2 \mathrm{d}T(r)\right] \\ &= \frac{1}{t^2}\int_0^t \dot{h}_1^2(r) \mathbb{E}\left[e^{-k\tau(r)} f^{-2}(X_r') \left|\int_0^{\tau(r)} R_s \mathrm{d}B_s\right|^2\right] \mathrm{d}r, \end{aligned}$$

where $R_s := \big(W_s^{-1}R(//_{0,s}\,\cdot\,, W_s v)W_s v\big)$. So far, these steps are quite similar to the proof of the lemma before.
Application of the Cauchy-Schwartz inequality to the expectation provides

$$\begin{aligned}
&\mathbb{E}\left[e^{-k\tau(r)}f^{-2}(X'_r)\left|\int_0^{\tau(r)} R_s \mathrm{d}B_s\right|^2\right] \\
&\leq \left(\mathbb{E}\left[e^{-2k\tau(r)}f^{-4}(X'_r)\right]\mathbb{E}\left[\left|\int_0^{\tau(r)} R_s \mathrm{d}B_s\right|^4\right]\right)^{\frac{1}{2}}.
\end{aligned}$$

For the first term on the right hand side Itô's formula says

$$\begin{aligned}
\mathrm{d}\left[e^{-2k\tau(r)}f^{-4}(X'_r(x))\right] &\overset{m}{=} -2ke^{-2k\tau(r)}f^{-2}(X'_r(x))\mathrm{d}r + e^{-2k\tau(r)}L'f^{-4}(X'_r(x))\mathrm{d}r \\
&\leq 2\hat{c}_2(f)e^{-2k\tau(r)}f^{-4}(X'_r(x)),
\end{aligned}$$

where $L'f^{-4} = f^2Lf^{-4} = \frac{1}{2}f^2\Delta f^{-4} = (10|\operatorname{grad} f|^2 - 2f\Delta f)f^{-4}$ due to Lemma 2.22. Hence, the last inequality holds for the choice of

$$\hat{c}_2(f) := \frac{1}{2}\sup_D\left(-2kf^2 + 10|\operatorname{grad} f|^2 - 2f\Delta f\right)_+.$$

Again by the now well-known Gronwall procedure (including stopping and the Fatou argument) we deduce

$$\mathbb{E}\left[e^{-2k\tau(r)}f^{-4}(X'_r)\right] \leq f^{-4}(x)e^{2\hat{c}_2(f)r}.$$

To estimate the remaining expectation, we fix an arbitrary orthonormal basis $(e_i)_{1\leq i\leq n}$ of T_xM and write R_s^{ij} for the corresponding matrix of the stochastic endomorphism $R_s : T_xM \to T_xM$ and B_s^j for the components of B_s with respect to the basis.
We have

$$\left|\int_0^{\tau(r)} R_s \mathrm{d}B_s\right|^4 = \left[\sum_i\left(\int_0^{\tau(r)}\sum_j R_s^{ij}\mathrm{d}B_s^j\right)^2\right]^2 \leq n\sum_i\left(\int_0^{\tau(r)}\sum_j R_s^{ij}\mathrm{d}B_s^j\right)^4$$

Now Burkholder-Davis-Gundy for $M^i := \int\sum_j R_s^{ij}\mathrm{d}B_s^j$ yields

$$\mathbb{E}[(M^i_{\tau(r)})^4] \leq C_4\mathbb{E}\left[\left(\int_0^{\tau(r)}\sum_j (R_s^{ij})^2\mathrm{d}s\right)^2\right] = C_4\mathbb{E}\left[\left(\int_0^{\tau(r)}(R_s^*R_s)^{ii}\,\mathrm{d}s\right)^2\right]$$

(for an explicit value of the constant C_4 see a proof of the upper Burkholder-Davis-Gundy inequality, e.g. [R-Y], Prop. IV.4.3).
The integrand satisfies $|(R_s^*R_s)^{ii}| \leq \|R_s\|^2 \leq \|W_s^{-1}\|^2\|W_s\|^4C_2^2$ with C_2 given as in the theorem.

Consequently, $\int_0^{\tau(r)}(R_s^*R_s)^{ii}\mathrm{d}s \le C_2^2\int_0^{\tau(r)} e^{(K-2k)s}\mathrm{d}s \le \frac{C_2^2}{K-2k}e^{(K-2k)\tau(r)}$ in the case of $K > 2k$. So we finally arrive at

$$\begin{aligned}\mathbb{E}\left|\int_0^{\tau(r)} R_s\mathrm{d}B_s\right|^4 &\le n^2C_4\left(\frac{C_2^2}{K-2k}\right)^2\mathbb{E}[e^{2(K-2k)\tau(r)}]\\ &\le n^2C_4\left(\frac{C_2^2}{K-2k}\right)^2 e^{2(K-2k)(\sup_D f^2)r}.\end{aligned}$$

For the last step we have used $\tau(r) = \int_0^r f^2(X'_s(x))\mathrm{d}s$.
The case $K < 2k$ is even easier since then $\int_0^{\tau(r)}(R_s^*R_s)^{ii}\mathrm{d}s \le \frac{C_2^2}{2k-K}$ and the estimate before holds with the exponential replaced by 1.
Taking into account the definition of $c_2(f) = \hat{c}_2(f) + (K-2k)(\sup_D f^2)$ for $K > 2k$ and $c_2(f) = \hat{c}_2(f)$ otherwise, cf. (4.34), we put our results into the first estimate of the present proof and find

$$\mathbb{E}[G^2_{\sigma\wedge\frac{t}{2}}] \le \frac{n\sqrt{C_4}C_2^2}{(K-2k)f^2(x)t^2}\int_0^t \dot{h}_1^2(r)e^{c_2(f)r}\mathrm{d}r.$$

We increase the right hand side by taking $c(f)$ instead of $c_2(f)$ and use the definition (4.43) of h_1 to end up with (4.49). □

Lemma 4.20. *For the processes I and J from (4.31) and (4.32) (with K_r and L_r given as in Proposition 4.16) holds*

$$\mathbb{E}\left[I^2_{\tau\wedge t}J^2_{\sigma\wedge\frac{t}{2}}\right] \le \frac{c(f)}{f^2(x)(1-e^{-c(f)t})}\frac{(1\vee e^{-k\frac{t}{2}})\tilde{c}(\tilde{f})}{\tilde{f}^2(x)(1-e^{-\tilde{c}(\tilde{f})t})}. \tag{4.50}$$

Proof. For brevity we will write $\mathscr{F}_r$ instead of $\mathscr{F}_r^{X(x)}$ throughout this proof.
We start out by $\mathbb{E}\left[I^2_{\tau\wedge t}J^2_{\sigma\wedge\frac{t}{2}}\right] = \mathbb{E}\left[\mathbb{E}\left[I^2_{\tau\wedge t}|\mathscr{F}_{\sigma\wedge\frac{t}{2}}\right]J^2_{\sigma\wedge\frac{t}{2}}\right]$.
For the same reasons as in the preceding proofs I is a $L^2(\mathbb{P})$-bounded martingale on $[0,\tau\wedge t]$ which equals 0 on $[0,\sigma\wedge\frac{t}{2}]$, hence we know that $I^2 - \int dIdI$ is a martingale vanishing up to time $\sigma\wedge\frac{t}{2}$. Consequently, we have $\mathbb{E}\left[I^2_{\tau\wedge t}|\mathscr{F}_{\sigma\wedge\frac{t}{2}}\right] = \mathbb{E}\left[\int_{\sigma\wedge\frac{t}{2}}^{\tau\wedge t} dIdI|\mathscr{F}_{\sigma\wedge\frac{t}{2}}\right]$ and due to the results of Proposition 4.16 follows

$$\begin{aligned}\mathbb{E}\left[I^2_{\tau\wedge t}\middle|\mathscr{F}_{\sigma\wedge\frac{t}{2}}\right] &= \mathbb{E}\left[\int_0^{\tau\wedge t}|//_{0,r}^{-1}W_r\dot{K}_r|^2\mathrm{d}r\middle|\mathscr{F}_{\sigma\wedge\frac{t}{2}}\right]\\ &= \frac{1}{t^2}\mathbb{E}\left[\int_0^{\tau\wedge t} e^{-kr}\dot{k}_r^2\mathrm{d}r\middle|\mathscr{F}_{\sigma\wedge\frac{t}{2}}\right]\\ &= \frac{1}{t^2}\mathbb{E}\left[\int_{\sigma\wedge\frac{t}{2}}^{\tilde{\tau}(t)} e^{-kr}\dot{\tilde{h}}_1^2(\tilde{h}_0(r))\tilde{f}^{-4}(X_r(x))\mathrm{d}r\middle|\mathscr{F}_{\sigma\wedge\frac{t}{2}}\right]\\ &= \frac{1}{t^2}\mathbb{E}\left[\int_{\sigma\wedge\frac{t}{2}}^{\tilde{\tau}(t)} e^{-kr}\dot{\tilde{h}}_1^2(\tilde{h}_0(r))\tilde{f}^{-2}(X_r(x))\mathrm{d}\widetilde{T}(r)\middle|\mathscr{F}_{\sigma\wedge\frac{t}{2}}\right]\\ &= \frac{1}{t^2}\mathbb{E}\left[\int_0^t \dot{\tilde{h}}_1^2(r)e^{-k\tilde{\tau}(r)}\tilde{f}^{-2}(\widetilde{X}_r(x))\mathrm{d}r\middle|\mathscr{F}_{\sigma\wedge\frac{t}{2}}\right]\end{aligned}$$

$$= \frac{1}{t^2}\int_0^t \dot{h}_1^2(r)\mathbb{E}\left[e^{-k\tilde{\tau}(r)}\tilde{f}^{-2}(\widetilde{X}_r(x))\middle|\mathscr{F}_{\sigma\wedge\frac{t}{2}}\right]\mathrm{d}r.$$

We reproduce the Gronwall argument from above for the conditional expectation. First of all, Itô's formula provides

$$\begin{aligned}\mathrm{d}\left(e^{-k\tilde{\tau}(r)}\tilde{f}^{-2}(\widetilde{X}_r(x))\right) = &-ke^{-k\tilde{\tau}(r)}\tilde{f}^{-2}(\widetilde{X}_r(x))\mathrm{d}\tilde{\tau}(r)\\ &-2e^{-k\tilde{\tau}(r)}(\tilde{f}^{-3}\operatorname{grad}\tilde{f})(\widetilde{X}_r)\mathrm{d}\widetilde{X}_r\\ &+e^{-k\tilde{\tau}(r)}\left(\tilde{f}^2\frac{1}{2}\Delta\tilde{f}^{-2}\right)(\widetilde{X}_r(x))\mathrm{d}r.\end{aligned}$$

Herein the integrated version of the second term on the right hand side stopped at suitable random times $\tilde{\sigma}_l := \inf\{s : f^{-2}\widetilde{X}_{\tilde{\tau}(s)} \geq l\}$ (cf. proof of 4.6) is a $\mathscr{F}_{\tilde{\tau}(r)}$-martingale. thus taking conditional expectation with respect to $\mathscr{F}_{\sigma\wedge\frac{t}{2}} = \mathscr{F}_{\tilde{\tau}(0)}$ yields its initial value 0. Since we have $\tilde{f}(\widetilde{X}_0(x)) = \tilde{f}(X_{\sigma\wedge\frac{t}{2}}(x)) = \tilde{f}(x)$ we find by integration over $[0,\sigma_l\wedge t]$, letting $l\to\infty$ and using Fatou

$$\begin{aligned}\mathbb{E}\left[e^{-k\tilde{\tau}(r)}\tilde{f}^{-2}(\widetilde{X}_r(x))\middle|\mathscr{F}_{\sigma\wedge\frac{t}{2}}\right] \leq &e^{-k(\sigma\wedge\frac{t}{2})}\tilde{f}^{-2}(x)\\ &+\int_0^r\tilde{c}(\tilde{f})\mathbb{E}\left[e^{-k\tilde{\tau}(s)}\tilde{f}^{-2}(\widetilde{X}_s(x))\middle|\mathscr{F}_{\sigma\wedge\frac{t}{2}}\right]\mathrm{d}s\end{aligned}$$

with $\tilde{c}(\tilde{f}) := \sup_D(-k\tilde{f}^2 + 3|\operatorname{grad}\tilde{f}|^2 - \tilde{f}\Delta\tilde{f})_+$. Thus, by the Gronwall lemma

$$\mathbb{E}\left[e^{-k\tilde{\tau}(r)}\tilde{f}^{-2}(\widetilde{X}_r(x))\middle|\mathscr{F}_{\sigma\wedge\frac{t}{2}}\right] \leq e^{-k(\sigma\wedge\frac{t}{2})}\tilde{f}^{-2}(x)e^{\tilde{c}(\tilde{f})r},$$

and consequently

$$\begin{aligned}\mathbb{E}\left[I^2_{\tau\wedge t}\middle|\mathscr{F}_{\sigma\wedge\frac{t}{2}}\right] &\leq \frac{e^{-k(\sigma\wedge\frac{t}{2})}}{t^2\tilde{f}^2(x)}\int_0^t\dot{h}_1^2(r)e^{\tilde{c}(\tilde{f})r}\mathrm{d}r\\ &= \frac{e^{-k(\sigma\wedge\frac{t}{2})}\tilde{c}(\tilde{f})}{\tilde{f}^2(x)(1-e^{-\tilde{c}(\tilde{f})t})}\,.\end{aligned}$$

A point which perhaps should be mentioned is the demand of the Gronwall reasoning for $\mathbb{E}\left[e^{-k\tilde{\tau}(r)}\tilde{f}^{-2}(\widetilde{X}_r(x))\middle|\mathscr{F}_{\sigma\wedge\frac{t}{2}}\right]$ to have a time-continuous version on $[0,t]$. This may be verified by the aid of the strong Markov property of the diffusion $\widetilde{X}_r$ which says that the conditional expectation factorizes over $\widetilde{X}_0(x) = X_{\sigma\wedge\frac{t}{2}}$, put into an expectation of a bounded measurable functional of the diffusion $X_r(x)$ itself. But this factor obviously is continuous with respect to time.

This result enables to estimate the expectation of I^2J^2. By the deterministic bound $e^{-k(\sigma\wedge\frac{t}{2})} \leq 1\vee e^{-k\frac{t}{2}}$ (which explodes at $t\to\infty$ if $k<0$), we continue with

$$\mathbb{E}[I^2_{\tau\wedge t}J^2_{\sigma\wedge\frac{t}{2}}] \leq \frac{(1\vee e^{-k\frac{t}{2}})\tilde{c}(\tilde{f})}{\tilde{f}^2(x)(1-e^{-\tilde{c}(\tilde{f})t})}\mathbb{E}[J^2_{\sigma\wedge\frac{t}{2}}]$$

With the same method as above we immediately find

$$\mathbb{E}[J^2_{\sigma\wedge\frac{t}{2}}] \le \frac{c(f)}{f^2(x)(1-e^{-c(f)t})}.$$

Combination of the last two estimates shows (4.50). □

The last proof shows that in the case of $k<0$, i.e. (the diagonal of) Ric takes negative values on D, there is a factor $e^{-kt/4}=e^{|k|t/4}$ in estimate (4.26) which explodes for $t\to\infty$. This term results from the bound on the norm of the damped transport W_s that increases up to time $\sigma\wedge\frac{t}{2}$. For this reason a corollary of Theorem 4.12 for harmonic functions by taking $t\to\infty$ to obtain a time-independent bound does only hold in the case of nonnegative Ricci curvature (at least on the smaller domain D).

Corollary 4.21. *Under the assumptions of Theorem 4.12 we consider $F_s=u$ to be bounded and harmonic on D and additionally* $\mathrm{Ric}(w,w)\ge 0$ *for all $w\in T_yM$, $y\in D$. Then we have the Hessian estimate*

$$|\operatorname{Hess}_x u| \le \frac{\|u\|_D\sqrt{c(f)}}{f(x)}\left(\frac{C_1}{|K-2k|}+\frac{C_2C_4^{1/4}\sqrt{n}}{\sqrt{|K-2k|}}+\frac{\sqrt{\tilde c(\tilde f)}}{\tilde f(x)}\right). \tag{4.51}$$

Proof. The only point to be clarified is that we just need $\|W_r\|\le 1$ on D to bound the explosive term of (4.26) by 1. Therefore we can relax the condition $k\ge 0$ to the weaker one on Ric asserted in the corollary. □

It remains to give explicit values for the constants appearing in Theorem 4.12 by suitable choices of the related domains and functions.
For this purpose we go back to the argument of Remark 4.11 instead of Theorem 4.9 itself since $\bar g''(0)=0$ for the function $\bar g$ from the remark. In contrast to this the cosine construction in the theorem (which just is C^1 at x) cannot be blown up to the outer part $B_R(x)\setminus B_{R/2}(x)$ of the geodesic ball with center x and radius R such that it remains C^2 when being continued constantly on $B_{R/2}(x)$.

Theorem 4.22. *Let $D\subset M$ be an open regular domain with $n=\dim(M)$. Consider the constants C_1, C_2, C_4, k and K from Theorem 4.12 except for the infimum of the Ricci curvature in (4.25) also taken on D.*
Then for any ($\frac12\Delta$-)space-time-harmonic function F on $[0,t]\times\bar D$ we have the following estimate for Hessian at $x\in D$:

$$\begin{aligned}
&|\operatorname{Hess}_x F_0| \le C_{\mathrm{Hess}}(r(x))\left(\|F_t\|_D \vee \sup_{[0,t]\times\partial D}|F|\right),\\
&C_{\mathrm{Hess}}(r) := \frac{\sqrt2\,C(r)}{\sqrt{1-e^{-C(r)^2t}}}\left(\frac{C_1}{|K-2k|}+\frac{C_2C_4^{1/4}\sqrt{n}}{\sqrt{|K-2k|}}+\frac{\sqrt2(1\vee e^{-kt/4})\widetilde C(r)}{\sqrt{1-e^{-\widetilde C(r)^2t}}}\right).
\end{aligned} \tag{4.52}$$

The constants depending on $r(x):=\operatorname{dist}(x,\partial D)$ are explicitly given by

$$C(r) := \sqrt{\left(\frac{6(n+16)}{r^2}+\frac{3\sqrt{(n-1)(-k)_+}}{r}+\frac{(-k)_+}{2}\right)_+ + \frac{(K-2k)_+}{2}} \tag{4.53}$$

and

$$\widetilde{C}(r) := \sqrt{\left(\frac{6(n+10)}{r^2} + \frac{3\sqrt{(n-1)(-k)_+}}{r} + \frac{(-k)_+}{2}\right)_+}. \tag{4.54}$$

Proof. As in the proof of 4.9 we first restrict ourselves to a ball in D with centre x on which the distance function is differentiable (except for x itself), thus let the radius of the ball be given by $\delta(x) := r(x) \wedge \operatorname{dist}(x, \operatorname{cut}(x))$.
To apply Theorem 4.12 for the choice of the two involved domains as $\widetilde{D} = B_{\delta(x)}(x)$ and $D = B_{\delta(x)/2}(x)$ we define

$$f(p) := \varphi(\operatorname{dist}(x,p)), \qquad \varphi(r) := \frac{1}{\sqrt{2}}\left[1 - \left(\frac{2r}{\delta(x)}\right)^3\right], \quad 0 \le r \le \frac{\delta(x)}{2},$$

and

$$\tilde{f}(p) := \psi(\operatorname{dist}(x,p)), \qquad \psi(r) := \begin{cases} \frac{1}{\sqrt{2}}, & 0 \le r < \frac{\delta(x)}{2}, \\ \frac{1}{\sqrt{2}}\left[1 - \left(\frac{2r}{\delta(x)} - 1\right)^3\right], & \frac{\delta(x)}{2} \le r \le \delta(x). \end{cases}$$

To clarify the geometrical situation, we mention that for $\operatorname{dist}(x,p) \ge \frac{\delta(x)}{2}$ the second definition can also be read as $\tilde{f}(p) = \frac{1}{\sqrt{2}}\left[1 - \left(\frac{2\operatorname{dist}(\partial B_{\delta(x)/2}, p)}{\delta(x)}\right)^3\right]$.
Note that in particular $\tilde{f}$ is C^2 at $\partial B_{\delta(x)/2}(x)$. We get the following derivatives of φ at $r \le \delta(x)/2$ and ψ at $\delta(x)/2 \le r \le \delta(x)$:

$$\varphi'(r) = -\frac{3\sqrt{2}}{\delta(x)}\left(\frac{2r}{\delta(x)}\right)^2, \qquad \varphi''(r) = -\frac{12\sqrt{2}}{\delta(x)^2}\left(\frac{2r}{\delta(x)}\right),$$

$$\psi'(r) = -\frac{3\sqrt{2}}{\delta(x)}\left(\frac{2r}{\delta(x)} - 1\right)^2, \qquad \psi''(r) = -\frac{12\sqrt{2}}{\delta(x)^2}\left(\frac{2r}{\delta(x)} - 1\right).$$

Herein the terms in large brackets are all bounded from above by 1. According to $|\operatorname{grad}\operatorname{dist}(x,\,\cdot\,)| \equiv 1$ we have

$$-\Delta f(p) = -\Delta(\varphi \circ \operatorname{dist}(x,\,\cdot\,))(p) = -\varphi'(\operatorname{dist}(x,p))\Delta\operatorname{dist}(x,\,\cdot\,)(p) - \varphi''(\operatorname{dist}(x,p))$$

and the same for $\tilde{f}$ with φ replaced by ψ. Just as in the gradient case comparison of the Laplacian of the distance function with the corresponding function of a model manifold of constant radial curvature $-\frac{(-k)_+}{n-1}$ provides

$$\Delta\operatorname{dist}(x,\,\cdot\,)(p) \le \sqrt{(n-1)(-k)_+} + \frac{n-1}{\operatorname{dist}(x,p)},$$

where we have used $\coth r \le 1 + \frac{1}{r}$ (cf. proof of Thm. 4.9). Moreover, $|\operatorname{grad} f|^2(p) \le \sup_{]0,\delta(x)/2[}(\varphi')^2$ and $|\operatorname{grad}\tilde{f}|^2(p) \le \sup_{]0,\delta(x)[}(\psi')^2$ This provides explicit upper bounds for the constants of the theorem, actually

$$c(f) \le \left(+\frac{(-k)_+}{2} + \frac{3\sqrt{(n-1)(-k)_+}}{\delta(x)} + \frac{6(n+16)}{\delta(x)^2}\right)_+ + \frac{(K-2k)_+}{2}$$

and
$$\tilde{c}(\tilde{f}) \le \left(+ \frac{(-k)_+}{2} + \frac{3\sqrt{(n-1)(-k)_+}}{\delta(x)} + \frac{6(n+10)}{\delta(x)^2} \right)_+ .$$
Hence the theorem is proven for $r(x)$ replaced by $\delta(x)$.
For the general case where $B_{r(x)}(x)$ may have nonempty intersection with $\mathrm{cut}(x)$ we carry over the argument of the proof of Theorem 4.9 based on Kendall's theorem. □

4.3. Pointwise gradient and Hessian estimates for positive harmonic functions on a rotationally symmetric manifold

In the general treatment of the foregoing parts of this chapter we applied the Cauchy-Schwartz inequality to factorize the expectation in our representation formulae and then provided estimates of the L^2-norms of the factors. The disadvantage of this method is that as the square of the real martingale $F_s \circ X_s(x)$ is a submartingale, its expectation increases in time, and consequently we obtained estimates for positive harmonic functions of the form
$$|\operatorname{grad}_x u| \le C\, \sqrt{u(x)\|u\|_D}$$
for a suitable constant C (and the same kind of result for the Hessian), which are local in that sense that they involve the supremum of u on a compact neighbourhood $\bar{D}$ of x.
Deterministic estimates on the gradient and Hessian of positive harmonic functions, however, can be found such that the right hand side only consists of the value $u(x)$ itself and a constant depending on curvature bounds and the distance from x to the boundary of the domain of harmonicity.

In this final paragraph we present a way to derive this kind of pointwise estimates of the first and second derivative of a positive harmonic function u at x in the particular case that our manifold M is rotationally symmetric with respect to x, which provides some sort of independence of the factors in the estimates.
The author owes the idea and the outline of the proof of this independence property to Marc Arnaudon.

Definition 4.23. A Riemannian manifold (M, g) is said to be *rotationally symmetric* with respect to $x \in M$, if for any (linear) isometry φ of the tangent space T_xM at x there exists an isometry $\theta : M \to M$ that leaves x fixed, such that $\varphi = (d\theta)_x$.

In the gradient case the following result holds:

Theorem 4.24. *Assume we have $x \in M$ such that M (of dimension $n \ge 2$ as in the previous parts of the chapter) is rotationally symmetric with respect to x and let $u : M \to \mathbb{R}_+$ denote a positive harmonic function on M.*
Let $D := B_R(x)$ denote the open geodesic ball of radius $R > 0$ around x. The Ricci curvature on D is assumed to be bounded from below by k, i.e.
$$k \le \inf_{y \in D,\, w \in T_yM,\, |w|=1} \operatorname{Ric}(w, w).$$

Then we have the pointwise gradient estimate

$$|\operatorname{grad}_x u| \le C_{\mathrm{grad}}(R)\, u(x) \tag{4.55}$$

with $C_{\mathrm{grad}}(R)$ *given by*

$$C_{\mathrm{grad}}(R) := \frac{1}{2}\sqrt{\frac{\pi^2(n+3)}{R^2} + \frac{2\pi\left(\sqrt{(-k)_+(n-1)}\right)}{R} + 4(-k)_+}. \tag{4.56}$$

The similar Hessian estimate is the following:

Theorem 4.25. *Let M (of dimension $n \ge 2$) be rotationally symmetric with respect to $x \in M$, $u : M \to \mathbb{R}_+$ positive harmonic and $t > 0$. Assume that for the domains $D = B_{R/2}(x)$ and $\widetilde{D} = B_R(x)$, $R > 0$, we are given $k, K \in R$ with*

$$k \le \inf_{y\in\widetilde{D}, w\in T_yM, |w|=1} \operatorname{Ric}(w,w) \le \sup_{y\in D,\, w\in T_yM,\, |w|=1} \operatorname{Ric}(w,w) \le K.$$

Then

$$|\operatorname{Hess}_x u| \le C_{\mathrm{Hess}}(R)\, u(x) \tag{4.57}$$

with

$$C_{\mathrm{Hess}}(R) := \frac{\sqrt{2}\, C(R)}{\sqrt{1-e^{-C(R)^2 t}}}\left(\frac{C_1}{|K-2k|} + \frac{C_2 C_4^{1/4}\sqrt{n}}{\sqrt{|K-2k|}} + \frac{\sqrt{2}(1\vee e^{-kt/4})\widetilde{C}(R)}{\sqrt{1-e^{-\widetilde{C}(R)^2 t}}}\right). \tag{4.58}$$

The constants C_1, C_2 and C_4 used herein are exactly the same as in Theorem 4.12, namely

$$C_1 := \sup_{y\in D,\, w\in T_yM, |w|=1} |(\nabla \operatorname{Ric}^\sharp + d^* R)(w,w)|$$

and

$$C_2 := \sup_{y\in D,\, w,w'\in T_yM,\, |w|=|w'|=1} |R(w,w')w'|.$$

C_4 denotes the upper constant in the Burkholder-Davis-Gundy inequality in the case $p = 4$. Finally, suitable constants $C(R)$ and $\widetilde{C}(R)$ are given by

$$C(R) := \sqrt{\left(\frac{6(n+16)}{R^2} + \frac{3\sqrt{(n-1)(-k)_+}}{R} + \frac{(-k)_+}{2}\right)_+ + \frac{(K-2k)_+}{2}}$$

and

$$\widetilde{C}(R) := \sqrt{\left(\frac{6(n+10)}{R^2} + \frac{3\sqrt{(n-1)(-k)_+}}{R} + \frac{(-k)_+}{2}\right)_+}.$$

Proof of Theorem 4.24. The main tool for the proof is the extension of the underlying probability space as follows.

Let $G := \{\theta : M \to M,\ \theta(x) = x,\ \theta \text{ isometry}\}$ the group of isometries on M that fix x. It is well-known that G is a Lie group isomorphic to the orthonormal group $O(n)$. On the Borel-σ-field $\mathcal{B}(G)$ of G consider the Haar measure μ of total mass 1 which is invariant

under left action of the group, i.e. $\mu(\theta(A)) = \mu(A)$ for all $A \in \mathcal{B}(G)$, $\theta \in G$ (for the construction of μ see for example [El 1], VIII.3).
Consequently, we consider the product space $(\Omega'; \mathscr{F}'; \mathbb{P}')$ given by

$$\Omega' := \Omega \times G, \qquad \mathscr{F}' := \mathscr{F} \otimes \mathcal{B}(G), \qquad \mathbb{P}' := \mathbb{P} \otimes \mu.$$

We claim that we obtain a M-valued Brownian motion Y (starting in x) on the product space by defining

$$Y_r(\omega, \theta) := \theta(X_r(x))(\omega).$$

The according filtration (satisfying the usual completeness and right continuity conditions) is derived by taking $\mathscr{F}'_r := \sigma\big(\mathscr{F}_r^{X(x)} \otimes \mathcal{B}(G), \mathcal{N}\big)$, where $\mathcal{N}$ denotes the σ-ideal of all sets in $\mathscr{F} \otimes \mathcal{B}(G)$ of outer measure 0.
We verify that Y is a Brownian motion on Ω' via stochastic development techniques and refer to [H-Th], p. 429-431, for details.
Consider a T_xM-valued Brownian motion Z_r (started at 0) which is adapted to $\big(\mathscr{F}_r^{X(x)}\big)$ and such that $X_r(x)$ is the stochastic development of Z_r.
Then we get a T_xM-valued process Z'_r on the product space Ω' by $Z'_r(\omega, \theta) := (d\theta)_x Z_r(\omega)$.
Obviously, Z'_r is a continuous process adapted to $(\mathscr{F}'_r)$. It is a martingale with respect to that filtration, since for all $A \in \mathscr{F}_r^{X(x)}$, $B \in \mathcal{B}(G)$, and $0 \le r \le s$ we compute

$$\begin{aligned}
\mathbb{E}[1_{A\times B} Z'_s] &= \int_G \mathrm{d}\mu(\theta) 1_B(\theta) \int_\Omega \mathrm{d}\mathbb{P}(\omega) 1_A(\omega) Z'_s(\omega, \theta) \\
&= \int_G \mathrm{d}\mu(\theta) 1_B(\theta) (d\theta)_x \int_\Omega \mathrm{d}\mathbb{P}(\omega) 1_A(\omega) Z_r(\omega) \\
&= \int_G \mathrm{d}\mu(\theta) 1_B(\theta) (d\theta)_x \int_\Omega \mathrm{d}\mathbb{P}(\omega) 1_A(\omega) Z_s(\omega) \\
&= \int_G \mathrm{d}\mu(\theta) 1_B(\theta) \int_\Omega \mathrm{d}\mathbb{P}(\omega) 1_A(\omega) Z'_s(\omega, \theta) \\
&= \mathbb{E}[1_{A\times B} Z'_r].
\end{aligned}$$

Finally, we observe that the quadratic covariation process of the components (with respect to a fixed orthonormal basis of T_xM) of Z' is the same as that according to the Brownian motion Z since the linear isometry $(d\theta)_x$ does not affect the inner product in the sums

$$\sum_{k=1}^{l-1} \Big\langle (d\theta)_x \left(Z^i_{t_{k+1}} - Z^i_{t_k} \right), (d\theta)_x \left(Z^j_{t_{k+1}} - Z^j_{t_k} \right) \Big\rangle_{T_xM}$$

and we obtain the bracket process by taking uniform limits in probability on compact time intervals.
For this reason Z'_r is a T_xM-valued $(\mathscr{F}'_r)$-adapted Brownian motion and we claim that Y_r is its stochastic development which shows that Y_r is a Brownian motion taking values in M.
This is proved if $\theta(X_r(x))$ is the stochastic development of $(d\theta)_x Z_r$ for all $\theta \in G$. But the latter fact is obvious when $X_r(x)$ is replaced by a C^1-curve $\gamma : [0,t] \to M$ with $\gamma(0) = x$ and we consider the Cartan development of a C^1-path in T_xM (which we identify with $\mathbb{R}^n$ by fixing an orthonormal basis) instead of the stochastic development (one could imagine

that θ and $(d\theta)_x$ yield an isometric transformation of the whole manifold M and the tangent space T_xM pinned onto M at x). But then as a general principle one can switch backwards from the curve solving an ordinary differential equation with respect to time and using the Cartan development to the semimartingale solving the according Stratonovich equation and the stochastic development, cf. [H-Th], p. 431.

Now the idea is to rewrite our intrinsic gradient Bismut formula in terms of the Brownian motion Y_r instead of $X_r(x)$ and emphasize the linear dependence of the involved finite energy process $K_r(v)$ of its initial value $v \in T_xM$. For this reason we now let K_r take values in $\mathrm{End}(T_xM)$ by defining

$$K_r(\,\cdot\,) := k_r \frac{1}{t}\, \mathrm{id}_{T_xM}$$

with the scalar process $k_r = h_1 \circ h_0(r)$ constructed as in Proposition 4.6. Herein h_1 is given by (4.17) and f only depends on the distance from x on $B_R(x)$; the explicit constant of the theorem is obtained by the choice of $f(p) = \cos\left(\frac{\pi}{2R}\, \mathrm{dist}(x,p)\right)$.
Clearly, $K_r(v)$ now is the same process as K_r in 4.6 was and we have the Bismut formula

$$\langle \mathrm{grad}_x\, u, v\rangle = \mathbb{E}\left[u(Y_{\tau\wedge t}) S_{\tau\wedge t}(v)\right]$$

with τ the first exit time of $Y_{\bullet}$ from $B_R(x)$.
The stochastic integral $S_{\tau\wedge t}(v)$ on the product space Ω' depending linearly on v in the Bismut formula is given by

$$S_{\tau\wedge t}(v) := \int_0^{\tau\wedge t} \langle W_r \dot{K}_r(v), \mathrm{d}^{\nabla} Y_r\rangle,$$

where one should notice that the Dohrn-Guerra transport W_r is taken along the paths $r \mapsto Y_r(\omega,\theta)$ of Y.
According to the rotational symmetry, $\tau(\omega,\theta)$ does not depend on θ, and its distribution with respect to ω equals the distribution of the first exit time of $X_{\bullet}(x)$ from $B_R(x)$ (which is the same as for $\theta \circ X_{\bullet}(x)$).
Then $S_{\tau\wedge t}$ has the rotational invariance property

$$S_{\tau\wedge t}(v)(\omega,\theta) = S_{\tau\wedge t}((d\theta)_x^{-1} v)(\omega, \mathrm{id}_M). \tag{4.59}$$

To show this equation, we need several consequences of the rotationally symmetric situation. So fix a $\theta \in G$, then first of all $\nabla d\theta = 0$, i.e. θ is affine, since an isometry maps geodesics onto geodesics (cf. [H-Th], 7.152, 7.155 Kor. 2 and 7.238). Hence $Y_r(\,\cdot\,,\theta) = \theta X_r(x)$ (for fixed θ) is again a Brownian motion on M satisfying the Itô-stochastic equation $\mathrm{d}^{\nabla}(\theta X_r(x)) = (d\theta)_{X_r(x)} \mathrm{d}^{\nabla} X_r(x) + 0 = (d\theta)_{X_r(x)} //_{0,r} \mathrm{d}B_r$.
Secondly, τ does not depend on θ, $\tau(\omega,\theta) = \tau(\omega, \mathrm{id}_M)$, according to the choice of τ as the first exit time of Y from the (rotationally symmetric) ball $B_R(x)$ (and this equals the first exit time of X from the ball for a.e. ω).
As third consequence, $(d\theta)$ commutes with curvature terms. In fact, for the curvature tensor R we have

$$(d\theta)_x R_x(u,v,w) = R_{\theta(x)}\left((d\theta)_x u, (d\theta)_x v, (d\theta)_x w\right)$$

and the corresponding results hold true for $\mathrm{Ric}^\sharp$, etc.
To the fourth and last, the damped transport $W_r(\,\cdot\,,\theta)$ along the paths of $\theta X_r(x)$ is given by

$$W_r(\,\cdot\,,\theta) = (d\theta)_{X_r(x)} W_r(\,\cdot\,,\mathrm{id}_M)(d\theta)_x^{-1}.$$

This is verified by noting that the right hand side is a transport along the paths of $\theta X_s(x)$ with initial value id_{T_xM} (both of these facts are clear) and satisfying the pathwise equation

$$DW_r(\,\cdot\,,\theta) = -\frac{1}{2}\,\mathrm{Ric}^\sharp_{\theta X_r(x)}\, W_r(\,\cdot\,,\theta)\mathrm{d}r.$$

Concerning the last property the product formula yields (according to $\nabla d\theta = 0$)

$$\begin{aligned}
DW_r(\,\cdot\,,\theta) &= 0 + (d\theta)_{X_r(x)} DW_r(\,\cdot\,,\mathrm{id}_M)(d\theta)_x^{-1}\mathrm{d}r \\
&= (d\theta)_{X_r(x)} \left(-\frac{1}{2}\,\mathrm{Ric}^\sharp_{X_r(x)}\, W_r(\,\cdot\,,\mathrm{id}_M)\right)(d\theta)_x^{-1}\mathrm{d}r \\
&= -\frac{1}{2}\,\mathrm{Ric}^\sharp_{\theta X_r(x)} (d\theta)_{X_r(x)} W_r(\,\cdot\,,\mathrm{id}_M)(d\theta)_x^{-1}\mathrm{d}r \\
&= -\frac{1}{2}\,\mathrm{Ric}^\sharp_{\theta X_r(x)}\, W_r(\,\cdot\,,\theta)\mathrm{d}r.
\end{aligned}$$

Prepared with these facts we can easily verify equation (4.59). One should note that since the real process k_r only depends on the distance from x, it is invariant under the isometry θ. Hence

$$\begin{aligned}
S_{\tau\wedge t}(v)(\,\cdot\,,\theta) &= \int_0^{\tau\wedge t} \left\langle W_r(\,\cdot\,,\theta)\frac{\dot{k}_r}{t}v, \mathrm{d}^\nabla(\theta X_r(x))\right\rangle_{\theta X_r(x)} \\
&= \int_0^{\tau\wedge t} \left\langle (d\theta)_{X_r(x)} W_r(\,\cdot\,,\mathrm{id}_M)(d\theta)_x^{-1}\frac{\dot{k}_r}{t}v, (d\theta)_{X_r(x)}//_{0,r}\mathrm{d}B_r\right\rangle_{\theta X_r(x)} \\
&= \int_0^{\tau\wedge t} \left\langle W_r(\,\cdot\,,\mathrm{id}_M)\frac{\dot{k}_r}{t}(d\theta)_x^{-1}v, //_{0,r}\mathrm{d}B_r\right\rangle_{X_r(x)} \\
&= S_{\tau\wedge t}((d\theta)_x^{-1}v)(\,\cdot\,,\mathrm{id}_M).
\end{aligned}$$

Now the crucial point of the proof is the immediate consequence of (4.59) that

$$\|S_{\tau\wedge t}\|(\omega,\theta) := \sup_{v\in T_xM,\,|v|=1} |S_{\tau\wedge t}(\omega,\theta)(v)|$$

does not depend on θ (in contrast to $S_{\tau\wedge t}$ itself, which is the reason why this independence property only works for the estimates and not for the representation formula).
Hence the gradient estimate reads

$$\begin{aligned}
|\operatorname{grad}_x u| &\le \mathbb{E}[u(Y_{\tau\wedge t})\|S_{\tau\wedge t}\|] \\
&= \int_\Omega \mathrm{d}\mathbb{P}(\omega)\int_G \mathrm{d}\mu(\theta)\|S_{\tau(\omega)\wedge t}\|(\omega)u(Y_{\tau(\omega)\wedge t})(\omega,\theta) \\
&= \int_\Omega \mathrm{d}\mathbb{P}(\omega)\|S_{\tau(\omega)\wedge t}\|(\omega)\int_G \mathrm{d}\mu(\theta)u\left(\theta X_{\tau(\omega)\wedge t}(x)(\omega)\right).
\end{aligned}$$

But now for fixed $\omega \in \Omega$, the inner integral equals $F(x)$ by the *boundary integral representation* or *mean value property* of harmonic functions: The law of $\theta X_{\tau(\omega)\wedge t}(x)(\omega)$ with respect to μ (ω fixed) is the uniform law on the geodesic sphere with centre x and radius $\operatorname{dist}(x, X_{\tau(\omega)\wedge t}(x)(\omega))$, according to the G-invariance (and normalization) of μ. This could for instance be shown by using the (stochastic) boundary integral representation property of u for an independent Brownian motion started at x and stopped when first hitting the sphere of radius $\operatorname{dist}(x, X_{\tau(\omega)\wedge t}(x)(\omega))$. The law of this stopped Brownian motion at time ∞ obviously equals the uniform law on the considered sphere.
For this reason we obtain

$$|\operatorname{grad}_x u| \leq \int_\Omega \mathrm{d}\mathbb{P}(\omega)\|S_{\tau(\omega)\wedge t}\|(\omega)F(x) = F(x)\mathbb{E}\left[\|S_{\tau\wedge t}\|\right].$$

According to the construction above, it is obvious that $\|S_{\tau\wedge t}\|$ is the same quantity that was used in our former Bismut estimate (without adjoining G to the underlying probability space), so for that reason it remains to return to the proof of Theorem 4.1 with the special choice of $K_r(v) = k_r v/t$ (k_r depending only on $\operatorname{dist}(x, X_r(x))$) and this time start out by

$$|\operatorname{grad}_x u| \leq F(x) \left\| \sup_{v\in T_xM,\,|v|=1} \left| \int_0^{\tau\wedge t} \langle W_r \dot{K}_r(v), //_{0,r}\mathrm{d}B_r\rangle \right| \right\|_{L^1(\mathbb{P})}.$$

Thus, we have to carry out the estimate of the L^1-norm of the stochastic integral instead of the L^2-norm (in the earlier case we estimated the L^2-norm only depending on $|v|$, so the supremum in the norm will make no difference).
To finish the proof we simply use $\|\cdot\|_{L^1(\mathbb{P})} \leq \|\cdot\|_{L^2(\mathbb{P})}$ and transfer the result of Theorem 4.9. □

The argument in this proof transfers with a few minor changes to the Hessian case, so we only give a very brief outline of this.

Proof of Theorem 4.24. The only difference to the gradient estimate is that the stochastic integral term S here has to be replaced by

$$Q(\omega,\theta)(v,w) := G_{\sigma\wedge\frac{t}{2}}(\omega,\theta)(v,w) + \frac{1}{2}H_{\sigma\wedge\frac{t}{2}}(\omega,\theta)(v,w) + I_{\sigma\wedge\frac{t}{2}}(\omega,\theta)(v,w)J_{\tau\wedge t}(\omega,\theta)(v,w)$$

for $v, w \in T_xM$ and G, H, I and J taken from (4.29) up to (4.32) with K and L again changed into endomorphisms on T_xM acting linearly on v and w.
One now has to verify that

$$Q(\omega,\theta)(v,w) = Q(\omega,\mathrm{id}_M)((d\theta)_x^{-1}v, (d\theta)_x^{-1}w)$$

which we are not going to show in detail. In fact, this is (for each of the the four stochastic integrals defining Q) the same argument as in the proof of (4.59) above. $d\theta$ commutes with $\nabla\operatorname{Ric}^\sharp$ and d^*R since this holds for the curvature tensor itself, and one can also write the inverse damped transport $W_r^{-1}(\,\cdot\,,\theta)$ (backwards along the paths of $X_r(x)$) in terms of $W_r^{-1}(\,\cdot\,,\mathrm{id}_M)$ by going back to the defining pathwise equation of W_r^{-1}.
Consequently $\|Q\|(\omega,\theta) := \sup_{v\in T_xM,\,|v|=1}|Q(\omega,\theta)(v,v)|$ is independent of θ, and thus the estimate for the Hessian reads

$$|\operatorname{Hess}_x u| \leq \mathbb{E}[u(Y_{\tau\wedge t})\|Q\|] \leq F(x)\mathbb{E}[\|Q\|].$$

By estimating the L^1-norm of $\|Q\|$ by its L^2-norm we can use the results of Theorem 4.22. $\square$

Bibliography

[A-Th 1] M. Arnaudon & A. Thalmaier. *Stability of stochastic differential equations in manifolds.* Séminaire de Probabilités, XXXII, 188-214. Lect. Notes in Math. **1686**. Berlin: Springer, 1998.

[A-Th 2] M. Arnaudon & A. Thalmaier. *Complete lifts of connections and stochastic Jacobi fields.* J. Math. Pures Appl. **77** (1998), 283-315.

[A-Th 3] M. Arnaudon & A. Thalmaier. *Bismut type differentiation of semigroups.* Prob. Theory and Math. Stat. (Vilnius, 1998), 23-32, VSP/TEV, Utrecht and Vilnius, 1999.

[A-Th 4] M. Arnaudon & A. Thalmaier. *Horizontal martingales in vector bundles.* Séminaire de Probabilités, XXXVI, 419-456. Lect. Notes in Math. **1801**. Berlin: Springer, 2002.

[Bi] J.M. Bismut. *Large deviations and the Malliavin calculus.* Progress in Mathematics **45**. Boston: Birkhäuser, 1984.

[Ch-Y] S.Y. Cheng & S.T. Yau. *Differential equations on Riemannian manifolds and their geometric applications.* Comm. Pure Appl. Math. **28** (1975), 333-354.

[D-Th] B. Driver & A. Thalmaier. *Heat kernel derivative formulas for vector bundles.* J. Funct. Anal. **183** (2001), 42-108.

[El 1] J. Elstrodt. *Maß- und Integrationstheorie.* Second edition. Berlin: Springer, 1999.

[El 2] K.D. Elworthy. *Stochastic differential equations on manifolds.* Cambridge: Cambridge University Press, 1982.

[El 3] K.D. Elworthy. *Stochastic flows on Riemannian manifolds.* In: M. Pinski, V. Wihstutz. *Diffusion processes and related problems in analysis, Vol. II.* Boston: Birkhäuser, 1992, 37-72.

[El 4] K.D. Elworthy. *Geometric aspects of diffusions on manifolds.* In: École d´Été de Probabilités de Saint-Flour, XV-XVII (1985-1987), 227-425. Lect. Notes in Math. **1362**. Berlin: Springer, 1988.

[E-LJ-L] K.D. Elworthy, Y. Le Jan & X.-M. Li. *On the geometry of diffusion operators and stochastic flows.* Lect. Notes in Math. **1720**. Berlin: Springer, 1999.

[E-L 1] K.D. Elworthy & X.-M. Li. *Formulae for the derivatives of heat semigroups.* J. Funct. Anal. **125** (1994), 252-286.

[E-L 2] K.D. Elworthy & X.-M. Li. *A class of integration by parts formulae in stochastic analysis.* In: N. Ikeda et al. (Eds.). *Itô´s Stochastic Calculus and Probability Theory.* Tokyo: Springer, 1996, 15-30.

[E-Y] K.D. Elworthy & M. Yor. *Conditional expectations for derivatives of certain stochastic flows.* Séminaire de Probabilités, XXVII, 159-172. Lect. Notes in Math. **1557**. Berlin: Springer, 1993.

[Em 1] M. Emery. *Stochastic calculus in manifolds.* Berlin: Springer, 1989.

[Em 2] M. Emery. *Martingales continues dans les variétés differentiables.* Lectures on probability theory and statistics (Saint-Flour, 1998), 1-84. Lect. Notes in Math. **1738**. Berlin: Springer, 2000.

[H-Th] W. Hackenbroch & A. Thalmaier. *Stochastische Analysis.* Stuttgart: Teubner, 1994.

[Hs 1] E. P. Hsu. *Estimates of derivatives of the heat kernel on a compact Riemannian manifold.* Proc. Amer. Math. Soc. **127** (1999), no.12, 3739-3744.

[Hs 2] E. P. Hsu. *Stochastic analysis on manifolds.* Providence, RI: AMS, 2002.

[I-MK] K. Itô & H.P. McKean. *Diffusion processes and their sample paths.* Reprint of the 1974 edition. New York: Springer, 1996.

[Jo] J. Jost. *Riemannian geometry and geometric analysis.* Second edition. Berlin: Springer, 1998.

[Ke] W.S. Kendall. *The radial part of Brownian motion on a manifold: a semimartingale property.* Ann. Probab. **15** (1987), 1491-1500.

[Kl] W. Klingenberg. *Riemannian geometry.* Berlin: de Gruyter, 1982.

[Ku] H. Kunita. *Stochastic flows and stochastic differential equations.* Cambridge: Cambridge University Press, 1990.

[Li] X.-M. Li. *Strong p-completeness of stochstic differential equations and the existence of smooth flows on noncompact manifolds.* Prob. Th. Rel. Fields **100** (1994), 485-511.

[L-Y] P. Li & S.T. Yau. *On the parabolic kernel of the Schrödinger operator.* Acta. Math. **156** (1986), 153-201.

[M-St] P. Malliavin & D.W. Stroock. *Short time behavior of the heat kernel and its logarithmic derivatives.* J. Diff. Geom. **44** (1996), 550-570.

[Na] J. Nash. *The imbedding problem for Riemannian manifolds.* Ann. of Math. **63** (1956), 20-63.

[No] J.R. Norris. *Path integral formulae for heat kernels and their derivatives.* Prob. Th. Rel. Fields **94** (1993), 525-541.

[Pe] P. Petersen. *Riemannian Geometry.* New York: Springer, 1998.

[Pi] J. Picard. *Gradient estimates for some diffusion semigroups.* Prob. Th. Rel. Fields **122** (2002), 593-612.

[R-Y] D. Revuz & M. Yor. *Continuous martingales and Brownian motion.* Third edition. Berlin: Springer, 1999.

[Sch] R. Schoen. *The effect of curvature on the behavior of harmonic functions and mappings.* In: R. Hardt & M. Wolf (Eds.). *Nonlinear partial differential equations in differential geometry.* IAS/Park City Math. Ser. 2, 129-184. Providence, RI: AMS, 1996.

[St 1] D.W. Stroock. *An estimate on the Hessian of the heat kernel.* In: N. Ikeda et al. (Eds.). *Itô´s Stochastic Calculus and Probability Theory.* Tokyo: Springer, 1996, 355-371.

[St 2] D.W. Stroock. *An introduction to the analysis of paths on a Riemannian manifold.* Providence, RI: AMS, 2000.

[St-T] D.W. Stroock & J. Turetsky. *Short time behaviour of logarithmic derivatives of the heat kernel.* Asian J. Math. **1**, No.1 (1997), 17-33.

[St-Z] D.W. Stroock & O. Zeitouni. *Variations on a theme by Bismut. Hommage à P.A. Meyer et J. Neveu.* Astérisque **236** (1996), 291-301.

[Th 1] A. Thalmaier. *On the differentiation of heat semigroups and Poisson integrals.* Stoch. Stoch. Rep. **61** (1997), 297-321.

[Th 2] A. Thalmaier. *Some remarks on the heat flow for functions and forms.* Electron. Comm. Probab.**3** (1998), 43-49.

[Th-W] A. Thalmaier & F.-Y. Wang. *Gradient estimates for harmonic functions on regular domains in Riemannian manifolds.* J. Funct. Anal. **155** (1998), 109-124.

[Yau] S.T. Yau. *Harmonic functions on complete Riemannian manifolds.* Comm. Pure Appl. Math. **28** (1975), 201-228.