Randomness and Symmetry in Quantum Control

Jan Nöller

Born April 8th 1996 in Haan, Germany

February 14, 2022

Master's Thesis Mathematics

First Supervisor: Prof. Dr. Massimiliano Gubinelli

Second Supervisor: Prof. Dr. Tommaso Calarco
(University of Cologne & Forschungszentrum Jülich)

Co-Advisors: Dr. Robert Zeier, Dr. Felix Motzoi (FZ Jülich)

INSTITUT FÜR ANGEWANDTE MATHEMATIK

MATHEMATISCH-NATURWISSENSCHAFTLICHE FAKULTÄT DER

RHEINISCHEN FRIEDRICH-WILHELMS-UNIVERSITÄT BONN

CONTENTS

ABSTRACT

This interdisciplinary thesis aims at applying probabilistic methods and tools used in the theory of invariant diffusion processes to the setting of compact Lie groups in order to solve and discuss problems related to randomised benchmarking procedures of near-term quantum devices. Specifically we show how one can use degenerate quantum control systems to generate sufficiently uniform random distributions on compact Lie groups, and discuss some examples.

The work is organised as follows. In the introduction we provide some background on the general context, randomised benchmarking methods and discuss our results with some simulations. Part I is dedicated to the necessary prerequisites regarding quantum systems and the structures we are interested in. Following that, in Part II we first review stochastic integration/differentiation in order to properly characterise stochastic processes that originate in randomly controlled quantum systems (Chapter 4) and subsequently introduce the associated probability semigroups and their generators in Chapter 5. Finally in Part III we combine the setting of Part I with the tools from Part II, and show how we can use quantum control systems to efficiently simulate the uniform distribution on compact Lie groups (Chapters 6,7). Since this thesis relies on results from probability theory and representation theory of Lie algebras alike, the Appendix features two sections dedicated to each of these topics in order to provide a compact overview of the important results and definitions.

ACKNOWLDEGEMENTS

First and foremost I want to thank all my supervisors who were involved in this project in one way or another. Massimilano Gubinelli - for his precise hints and suggestions, Tommaso Calarco for accepting the second supervision of this thesis and in particular Robert Zeier and Felix Motzoi for many, many insightful discussions, and their constant support in helping me not to get lost in the quantum world. Also I want to thank Jorge Y. Malo and Johannes Kombe for providing more physical motivated context and their explanations in several sessions. Lastly I wish to express my overwhelming gratitude towards my parents who are always supportive in every imaginable way, and not last for sustaining me over the past week. The author acknowledges funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 817482 (PASQuanS).

INTRODUCTION AND SUMMARY

Motivation

Quantum information and quantum computing have the potential to create a significant impact in various fields of science and technology in the near term future, ranging from secure communication protocols to scientific applications in simulating quantum systems and molecule structures. The idea to use the laws of quantum mechanics for computational purposes was first suggested by Richard Feynman in the 1980's, who recognised that classical computers cannot simulate multi-particle quantum systems efficiently [1]. Since then a lot of effort has been put into the pursuit of the practical realisation of quantum computers. One major contribution suggesting the superiority of quantum computers for specific tasks is due Peter Shor, who devised a quantum algorithm that allows to efficiently factorise large integers into their primes [2], thereby undermining the premise of secure communication protocols based on RSA-cryptography. Despite considerable efforts in research, the quantum processors available today and in the near future only belong to the category of Noisy Intermediate Scale Quantum (NISQ) processors. These are characterised by a limited qubit range, and more importantly a lack of built-in procedures for error correction and are therefore sensitive to all kinds of noise. Fault-tolerant quantum computing requires much more resources than are available today, demanding possibly well over a hundred physical qubits in order to encode one single logical qubit [3]. The Sycamore quantum processor presented by Google researchers in 2019 only features 53 fully entangled qubits, and even the current leader, IBMs 'Eagle' processor consists merely of 127 qubits. In order to perform high-end computational tasks such as Shor's factorisation algorithm reliably, several more breakthroughs are needed in order to overcome the obstacles in scaling the underlying quantum systems.

Even though fault-tolerant quantum computers are currently out of reach, NISQ devices still promise to bring advantages in various fields of applications. Some of the most prominent ones are quantum simulations [4] and several classes of hybrid quantum-classical optimisation procedures, such as Quantum Approximate Opimisation Algorithms (QAOA) [5, 6, 7], Variational Quantum Algorithms (VQA)[8] and Variational Quantum Eigensolvers (VQE) [9]. The core idea of such algorithms is to relate an objective function to be minimised to the Hamiltonian of a quantum system. Then one can determine the ground state of the system by a combination of quantum simulation and classical optimising procedures on involved parameters. Several combinatorial

optimisation tasks related to graphs (e.g. MaxCut) have been investigated in this regard [10],[6]. One way or the other, one will have to settle with the fact that near term quantum devices should be treated with some caveats as far as the fidelity of the outcome is concerned. This makes it a necessity to reliably assess how accurate a quantum device of interest works.

Randomised Benchmarking

Errors that arise in quantum computations can be split into two major groups. First, we have errors that are related to state preperation and measurment (SPAM) and therefore *independent* of the computation time. On the other hand, there are errors which depend on the computation length, such as the noisy implementation of the gates themselves and leakage effects which are caused by the system not being properly isolated.

One possible way to proceed is by performing quantum state tomography. By performing different measurements on ensembles of states that have undergone a noisy evolution of a unitary U, we can reconstruct the outcome, and compare this with the state which we would expect if the gate implementation was error free. Repeating this for different quantum gates, we can create a map, telling us which operations are 'good' or 'bad'. This method however has several drawbacks. In order to properly determine how the the operation acts on one state using different measurements, we repreatedly need to obtain the same output state under the same input. Therefore this procedure is particularly sensitive to SPAM errors and also unreliable because the (non-deterministic) noisy gate implementation does not produce identical states which is required for the tomography. It is needless to say that it also scales rather badly in system size. Nevertheless, efficient and robust methods to estimate the fidelity of fixed gate sequences exist, as discussed, e.g. in [11]. But even here, handling the amount of data we retrieve for many different combinations of gates over the course of a computation becomes somewhat impractical. The natural solution is to consider *averaged* gate fidelities, leading to the various flavours of *randomised benchmarking*, which were first proposed in [12]. The key idea here is to first apply random gate sequences of fixed length to an initial state ρ and compare the results with the expected outcome. By varying the length of the sequence we can the efficiently extract the time-dependent error rates and the influence of SPAM errors. Most protocols for randomised benchmarking work along the lines of the algorithm in Tab.1, which was proposed by [13] and designed to be applicable in analogue quantum computing.

Randomised benchmarking procedures have several benefits. Besides providing a single figure of merit characterising the accuracy of the quantum processor, they are robust to noise and scalable [14, 15]. In order to produce reliable results and evenly detect the error rates, we are now faced with a different problem: We need to simulate distributions over unitary groups, that are sufficiently close to being unifom.

Randomised Benchmarking (analogue setting, as suggested in [13])

1. Randomly choose uniformly distributed unitaries $U_1, \ldots, U_n \in \mathrm{SU}(d)$ independently.
2. Repeatedly apply the sequence $S = [\Lambda_{U_1}, \Lambda_{U_1^\dagger}, \ldots, \Lambda_{U_n}, \Lambda_{U_n^\dagger}]$ to initial state ρ_0, where Λ_{U_i} denotes the channel representing the noisy implementation of the gate U_i. Compare the final state ρ_S with the input state ρ_0 and determine the survival probability $\mathrm{Tr}(\rho_S \rho_0^\dagger)$ by repeated measurements. If the gate implementation was perfect, this would return $\mathrm{Tr}(\rho_S \rho_0^\dagger) = 1$.
3. Repeat the two prior steps for various random sequences of the same length to obtain an average fidelity F_n for the computation length n
4. Repeat the steps above to determine F_n for different n.
5. Determine the constants A, B, f such that $F_n \approx A + Bf^n$. The constants A and B absorb SPAM errors, while $r := \frac{d-1}{d}(1-f)$ is the *average error rate*, the quantity we are ulimately interested in.

Tab. 1: Algorithm

In the setting of *discrete* quantum computing, this is accomlished by so called *unitary k-designs*,[1] which are finite subsets $D \subset \mathrm{SU}(d)$ satisfying

$$\frac{1}{|D|} \sum_{U \in D} U^{\otimes k} \otimes \bar{U}^{\otimes k} = \int_{\mathrm{SU}(d)} U^{\otimes k} \otimes \bar{U}^{\ tk} d\mu_{Haar}(U) \tag{1}$$

The moment operator of the Haar measure $\rho \mapsto \int_G U^{\otimes k} \rho U^{\dagger \otimes k} d\mu_{Haar}(U)$ is also known as the *twirling channel* in the physics literature.

Notation: In the following we will usually write $\Phi^{k,k}$ for the representation $U \mapsto U^{\otimes k} \otimes \bar{U}^{\otimes k} \in \mathrm{End}(\mathbb{C}^d)^{\otimes 2}$ and sometimes $M_\nu^k := \int_G \Phi_U^k d\nu(U)$ for the $\Phi^{k,k}$-moment of a probability distribution ν on $G \subset \mathrm{SU}(d)$.

Several gate sets have been identified to form 2-designs on $\mathrm{SU}(d)$, most notably the Clifford group, which even forms a 3-design [16]. One can also relax the requirement (1) only to hold up to an error in operator norm, leading to the notion of *approximate* designs. For instance fixed length products of universal gate sets (Def. 1.3) form approximate deigns of every order and arbitrarily small error, if the length is chosen to be sufficiently large [17]. Going back to the randomised benchmarking protocol, with such an (approximate) design $D \subset \mathrm{U}(d)$, we replace the uniform sampling on $\mathrm{SU}(d)$ in step 1 by averaging over the elements in D instead.

[1] In the literature, these are usually called t-designs. However confusing this with the time parameter at some stage is almost guaranteed, therefore we shall stick to k-designs for the remainder of this thesis.

However this only partially solves the original problem. After having theoretically identified some gate set to be an (approximate) design, it is generally neither easy nor efficient to physically implement them one by one on the quantum processor. It seems somewhat redundant to invest a lot of effort in implementing gates more or less precisely, which are supposed to simulate a close to uniform but nevertheless *random* distribution. This holds even more so if we are trying to benchmark quantum devices that are used for analogue simulation tasks. Since every quantum operation will eventually be physically realised using a quantum control system in one form or another, it is natural to switch to the underlying analogue setting and generate random distributions by applying random pulses to the control parameters we have over the system. Ideally, we would expect that after a sufficiently long mixing time, we obtain a distribution ν which is close enough to the Haar measure for our purposes. Therefore, we generalise the notion of unitary designs in the obvious way by replacing the averaged sum in the left hand side of (1) by an integration over the considered distribution ν on $\mathrm{SU}(d)$. Of course this means that we effectively give up the design-property, in the sense that designs are finite sets by definition. In the literature this aspect is largely ignored, we however shall refer to such distributions as (ε-approximate) k-*Haar measures* (Chapter 3).

The approach to generate random distributions on $\mathrm{SU}(d)$ using quantum control systems has been investigated for instance by Onorati et al. [18] and Banchi et al. [19]. In the scenario of [18], the authors proved that one needs to wait a time of $T(\varepsilon, k, d) \in \mathcal{O}(\log(d), \varepsilon, poly(k))$ in order to get ε-approximate k-Haar mesures on $\mathrm{SU}(d)$ (see Theorem 10 therein). However this is accomplished under the rather restrictive assumption, that one control term is available for every possible direction on the Lie group. While this might work well for smaller systems, such assumptions are usually beyond the scope of what is practically realisable as soon as the system scales. For the degenerate scenario on $\mathrm{SU}(d)$ - i.e. we have less controls than possible directions on the generated Lie group - the same qualitative result of exponential convergence still holds as argued in [19]. Due to a lack of control over the degeneracies, universally applicable convergence rates cannot be expected. Deriving concrete bounds therefore requires a detailed special case analysis, as for instance done by [19] for random walks on spin chains.

This thesis mainly evolves around the outline presented in [19], understanding and formalising the proof and exploring possible improvements. In particular, we manage to lift the restriction on $G = \mathrm{SU}(d)$ of [19] and can include all cases, in which the generated Lie group is *semisimple*. If the center is non-trivial, this still works as long as one chooses the drift and controls of the system carefully enough. The precise statement is summarised in Result 1. This more general scenario is relevant for near term applications, since in many cases one does not have control over the full available Hilbert space, as in the examples of [13] which we also used for our simulations. We still cannot derive easy computable bounds on the convergence rates. At the core we are facing the problem to find bounds on the convergence rate of so called *hypoelliptic diffusion processes*. Although such methods are discussed in several publications, [20, 21, 22, 23, 24] rooting in the

initial work by Bakry and Emery [25], several attempts to apply the available methods to the setup we are interested in have turned out to be ultimately unsuccessful. Nevertheless, we have the following two results, which clarify the conditions needed for convergence in all moments.

Results

Result 1 (Theorem 7.9)**.** Consider a control system with drift Hamiltonian H_0 and control Hamiltonians $H_1^c, \dots, H_n^c$ which generates a subgroup $G \subset \mathrm{SU}(d)$. Further we assume that the *zero-time ideal* $\mathfrak{I}_0$ of the control system satisfies $\mathfrak{I}_0 = \mathfrak{g}$. This conditions means that already the controls $iH_1^c, \dots, iH_n^c$ and the commutators $[iH_0, iH_1^c], \dots, [iH_0, iH_n^c]$ generate the Lie algebra of G, and ensures that the drift direction can be influenced by the control terms. In the following we shall call control system exhibiting these properties *regular*. Consider the random evolution on G determined by

$$\frac{d}{dt}U_t = -\Big(iH_0 + i\sum_j c_j(t)H_j^c\Big)U_t \tag{2}$$

where the c_j are piecewise constant, and normally distributed $\sim \mathcal{N}(0, \Delta T^{-1})$ for a sufficiently small time step $\Delta T > 0$. Then for all $k \in \mathbb{N}$ we have that

$$\lim_{t\to\infty} \mathbb{E}[U_t^{\otimes k} \otimes \overline{U}_t^{\otimes k}] = \int_G U^{\otimes k} \otimes \overline{U}^{\otimes k} d\mu_G(U).$$

The convergence rate is exponential and determined by the gap of the Lindblad operator $\mathcal{L}_k = i[H_0^{\otimes k}, \cdot] - \frac{1}{2}\sum_j [(H_j^c)^{\otimes k}, [(H_j^c)^{\otimes k}, \cdot]]$, where the tensor product is to be understood in the Lie algebra sense (see Ex. A.16).

The second result is essentially a corollary to Result 1 and Thm. 2.13 of [26].

Result 2 (Prop. 7.13)**.** On a compact *semisimple* Lie group G, the convergence of the process (2) to the Haar measure in all higher moments is completely encoded in the convergence of the second moment. Specifically, the following statements are equivalent:

1. $\lim_{t\to\infty} \mathbb{E}[U_t^{\otimes 2} \otimes \overline{U}_t^{\otimes 2}] = \mathbb{E}_{\mu_G}[U^{\otimes 2} \otimes \overline{U}^{\otimes 2}]$,
2. $\lim_{t\to\infty} \mathbb{E}[U_t^{\otimes k} \otimes \overline{U}_t^{\otimes k}] = \mathbb{E}_{\mu_G}[U^{\otimes k} \otimes \overline{U}^{\otimes k}]$ for all $k \in \mathbb{N}$.

The implications of these results are two-fold. First of all, Result A shows that it is possible to simulate uniform distributions on semisimple Lie groups even by the means of highly degenerate control systems. Secondly, already the second moment contains the information about the convergence to the uniform distribution. For the exact convergence speed, however one needs to take the convergence in moments with respect to all (irreducible) representations into acount again, which is of course not feasible for pracical purposes. In many applications having a 2-design is already completely suffcient. Consider again for instance the randomised enchmarking protocol 1,

where we want to estimate the average gate fidelity in steps 1. and 2. The survival probability of the initial state under the noisy implementation of a gate $U \in \mathrm{SU}(d)$ is given by

$$\mathrm{Tr}(\rho_0^\dagger \rho_S) = \mathrm{Tr}\left(\rho_0^\dagger \Lambda_{U^\dagger} \Lambda_U(\rho_0)\right). \tag{3}$$

For a linear noise model where we can express the noisy implementation with suitable A_i, B_i via

$$\Lambda_U(\rho) = \left(\sum_i A_i U B_i\right) \rho \left(\sum_i B_i^\dagger U^\dagger A_i^\dagger\right)$$

for all $U \in \mathrm{SU}(d)$, we can replace the integration of (3) over the Haar measure by averaging over a unitary 2-design on $\mathrm{SU}(d)$. This is also applicable to longer sequences when the random gates are chosen independently. In order to include higher orders in the noise model as well, one needs to consider designs of an appropriate higher degree instead.

Examples and Simulations

When applying Result 1 to randomised benchmarking procedures of analogue quantum devices, we will only take into account finitely many samples from this random distribution, and therefore arrive again at approximate (discrete) unitary k-designs. Assume that we have a measure ν on a compact $G \subset \mathrm{SU}(d)$ which is close to the Haar measure in the $\Phi^{k,k}$-moment. How uniform will the discrete sampling from this distribution turn out eventually? Or phrased differently, for independently chosen $U_1, \ldots, U_N \sim \nu$ how large is the expected difference of $\frac{1}{N}\sum_{i=1}^N \Phi_{U_i}^{k,k} - M_G^k$ e.g. in the the 2-norm? As a matter of fact, we can explicitly state the expected error by[2]

$$\mathbb{E}_{\nu^{\otimes n}}\left[\left\|\frac{1}{N}\sum_{i=1}^N \Phi_{U_i}^{k,k} - M_G^k\right\|_2^2\right] = \frac{d^{2k}}{N} + \frac{N^2-N}{N^2}\|M_\nu^k\|_2^2 - \|M_G^k\|_2^2. \tag{4}$$

For large N, we can put $\frac{N^2-N}{N^2} \approx 1$ and therefore $\frac{N^2-N}{N^2}\|M_\nu^k\|_2^2 - \|M_G^k\|_2^2 \approx \|M_\nu^k - M_G^k\|_2^2$, cf. (8).

In the following we consider systems of $2, 3$ and 4 qubits subject to an XY-Heisenberg interaction with a transverse magnetical field. This translates to the drift being described by the Hamiltonian

$$H_0 = \sum_{i<j} J_{ij}\left(\sigma_x^i \sigma_x^j + \sigma_y^i \sigma_y^j\right) + \sum_j \sigma_z^j, \tag{5}$$

where the coupling strength is captured by a real matrix (J_{ij}). We assume the control Hamiltonians of the system to be the terms $\sigma_x^i \sigma_x^j$ for which $J_{ij} \neq 0$. This setup is just one possible incarnation

[2] The computation is somewhat similar to the ones of Rem. 3.7 and in the proof of Thm. 3.4-2.

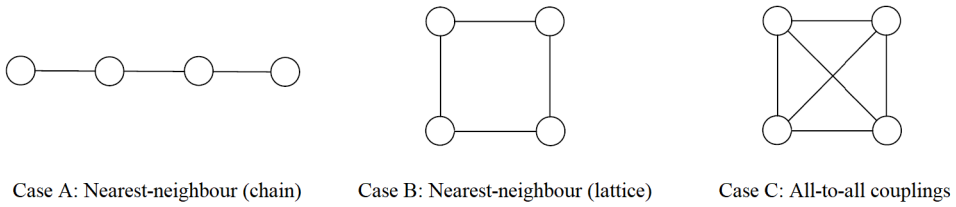

Fig. 1: Different interaction graphs representing J

of the Hubbard model which is ubiquitous in solid states physics. Potential platforms for quantum simulation based on this are for instance considered in [27] (without magnetic field) or in [13].
For our examples we assume the coupling strengths to be either 0 or 1. With different choices for (J_{ij}), we can capture different architectures such as nearest neighbour interactions ($J_{i,i+1} = 1$ and 0 otherwise), or all-to-all couplings ($J_{ij} = 1$ for all i, j). In the four-qubit scenarios A-C, we consider three different choices of coupling structures which are illustrated by the interaction graphs in Fig. 1. For the two- and three-qubit examples D,E we choose nearest-neighbour interactions on a chain, similar to A. The generated Lie algebras, dimensions and symmetries etc. in each of these cases are summarised in Fig. 2. We model the process for the respective choices of control systems via the random differential equation

$$\frac{d}{dt}U_t = -\Big(iH_0 + i\sum_j c_j(t)H_j^c\Big)U_t,\ U_0 = 1_G \tag{6}$$

where the c_j are piecewise constant and normally distributed $\sim \mathcal{N}(0, \Delta T^{-1})$ for the time step $\Delta T := \pi \times 10^{-3}$. We introduce the factor π because it takes the drift Hamiltonian the time $T = 2\pi$ to perform a full rotation on each of the Lie groups.
In order to quantify the difference to the uniform distribution that we get by discretely sampling from $\mu_t = Law(U_t)$ we use the *frame potential* of the sampled distribution. Recall that for a finite

	Lie algebra $\mathfrak{g}$	$\dim(\mathfrak{g})$	$\mathcal{F}_{\min}$	Full Lie algebra
A	$\mathfrak{so}_8$	28	18	$\mathfrak{su}_{16}$
B	$\mathfrak{so}_8 \oplus \mathfrak{so}_8$	56	12	$\mathfrak{su}_{16}$
C	$\mathfrak{su}_8 \oplus \mathfrak{su}_8$	126	8	$\mathfrak{su}_{16}$
D	$\mathfrak{su}_2 \oplus \mathbb{R}$	4	20	$\mathfrak{su}_4$
E	$\mathfrak{su}_4$	15	14	$\mathfrak{su}_8$

Fig. 2: The different Lie algebras, dimensions, minimum frame potential $\mathcal{F}_{\min}$ on the corresponding Lie group determined by the quadractic symmetries of the $\Phi^{2,2}$ representation, and the enclosing Lie algebra. The Lie algebras in A-C,E are semisimple and therefore automatically satisfy the assumption of Result 1 (see Remark 2.9). For the case D this is explicitly checked in Ex. 2.12. The Lie algebras and the respective symmetries were computed using the Magma computational algebra system [28].

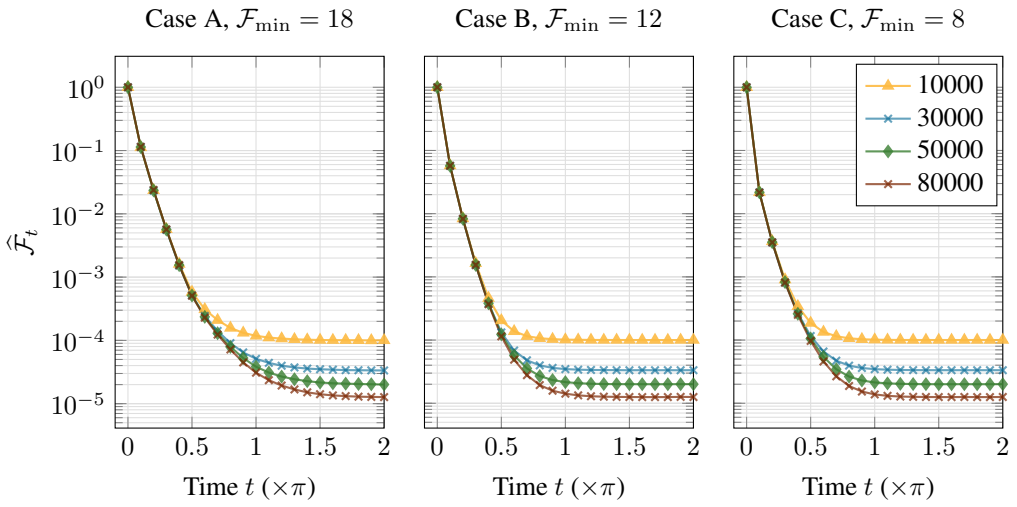

Fig. 3: Explanation: Reduced frame potential $\widehat{\mathcal{F}}_t$ as defined in (9) of $\{U_1, \dots, U_N\}$ for $U_1, \dots, U_N \sim \mu_t$ independent is plotted for different sample sizes N against the evolution time t in the respective cases with $\sigma_x^i\sigma_x^j$-controls and Hamiltonian (5) corresponding to the interaction graphs in Fig. 1. The time-step in the simulations is $\delta = 0.001(\times\pi)$, with 10 computed values of $\widehat{\mathcal{F}}_t$ per time unit. The Julia code can be found in Appendix C.

subset $S \subset G \subset \mathrm{SU}(d)$, the frame potential is defined by (cf. e.g. [29])

$$\mathcal{F}(S) := \frac{1}{|S|^2} \sum_{U,V \in S} |\operatorname{Tr}(U^\dagger V)|^4, \tag{7}$$

which coincides with the 2-norm of the moment M_S^2 of the uniform distribution over S with respect to the representation $\Phi^{2,2}$, see Rem. 3.7. The symmetries of the representation $\Phi^{2,2}$ then determine the minimum value $\mathcal{F}_{\min}$ the frame potential of any subset $S \subset G$ can attain, since

$$0 \leq \|M_S^\Phi - M_G^\Phi\|_2^2 = \|M_S^\Phi\|_2^2 - \|M_G^\Phi\|_2^2 \tag{8}$$

holds for all representations Φ (see proof of Thm. 3.4). For a better comparability of the scenarios with different system sizes, we add a factor d^{-4} to (7), where $\mathfrak{su}_d$ is the full enclosing Lie algebra. This amounts to the normalisation $\frac{1}{d}\operatorname{Tr}(\mathbb{1}_d) = 1$ of the trace. We therefore introduce the *reduced frame potential* for $S \subset G \subset \mathrm{SU}(d)$ by

$$\widehat{\mathcal{F}}(S) := \frac{1}{d^4}\Big(\mathcal{F}(S) - \mathcal{F}_{\min}(G)\Big), \tag{9}$$

which also absorbs the value of $\mathcal{F}_{\min}$. The plots in Fig. 3 and Fig. 4 show how $\widehat{\mathcal{F}}_t$ decays for different choices of control systems and sample sizes. In Fig. 4 we compare this also to the exact moment by computing the reduced 2-norm $d^{-4}\left(\|\exp(t\mathcal{L}_2)\|_2^2 - \mathcal{F}_{\min}\right)$ of the exponential

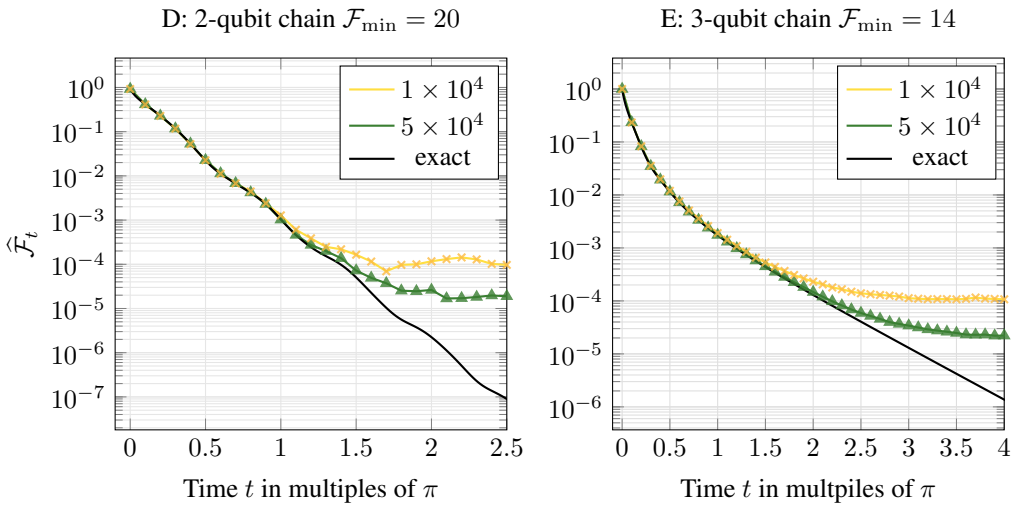

Fig. 4: Reduced frame potential $\widehat{\mathcal{F}}_t$ as in (9) of $\{U_1, \dots, U_N\}$ for $U_1, \dots, U_N \sim \mu_t$ independent, plotted for different sample sizes $N = 5 \times 10^4$ and $N = 1 \times 10^4$ against the evolution time t for the two cases D (2-qubit chain) and E (3-qubit chain). The time step in the simulations is $\delta = 0.001(\times\pi)$, with 10 computed values of $\widehat{\mathcal{F}}_t$ per time unit. The black line plots the corresponding reduced squared 2-norm of the exact moment operator, computed using the representation $M^2_{\mu_t} = \exp(t\mathcal{L}_2)$, cf. Prop. 7.2.

describing the moment semigroup in terms of its generator by $M^2_{\mu_t} = \exp(t\mathcal{L}_2)$ (see Prop. 7.2). In order to evaluate this operator exponential reliably, we need to work with exact arithmetics and therefore only compute this for these lower-dimensional examples. In each of the plots in Fig. 3 and Fig. 4, we see that initially the convergence is exponentially fast and independent of the sample size. Fig. 4 suggests that we more or less exactly recover the exact moment during this stretch of time. Depending on the number of samples, the decay then goes into a short transition and becomes stagnant after a while. Since the bound should become strictly better according to Eq. (7.18), this indicates that any improvement on the 'uniformity' of the distribution is subsequently lost in the error caused by discrete sampling.

We also observe that the level, on which $\widehat{\mathcal{F}}_t$ stabilises only depends on the chosen sample size and is approximately $1/N$ in each of the plots. This matches the predictions from Eq.(4) which we would expect when sampling from a k-Haar measure, after getting rid of the factor $d^{2k} = d^4$. These findings suggest that with the chosen number of samples and polynomials of bidegree $(2, 2)$, we cannot statistically distinguish the simulated distribution from the Haar measure anymore and therefore have reached the 'mixing time' in this sense. In the examples of Fig. 3 the mixing times for $N = 10^4, N = 5 \times 10^4$ are relatively short when compared to Fig. 4 and the exponential decay in Fig. 3 A,B,C is already stopped at $T \approx 0.5$ as opposed to $T \approx 1$ in Fig. 4 D and $T \approx 1.5$ in Fig 4 E respectively. The reason is that a fixed number of samples becomes less and less adequate to properly simulate the Haar measure with growing system sitze. A unitary 2-designs on $\mathrm{SU}(d)$ for instance needs to consist of at least $d^4 - 2d^2 + 2$ elements [29]. In this context it also needs

to be clarified what choice of norm and scale is suited to characterise the margin of error we want to allow for the purpose of randomised benchmarking. Several candidates are for instance discussed in Lemma 2.2.14 in [30], one of which is the *diamond norm* $\|\cdot\|_\diamond$ (see Def. 1.2). The diamond norm is frequently used in quantum information to characterise the distance between quantum channels and is the norm usually chosen to define approximate designs [19, 18, 30]. Since universal estimates relating $\|\cdot\|_2$ and $\|\cdot\|_\diamond$ are possibly overly pessimistic in general, one should probably investigate the following question: Given quantum channel Λ in dimension $d \times d$ which has 2-norm $\|\Lambda\|_2 = 1$, what is the expectation of $\|\Lambda\|_\diamond$, assuming a uniform distribution over these normalised quantum channels? Even if this does not give a definite answer on how the norms $\|M_{\mu_t}^k\|_2$ and $\|M_{\mu_t}^k\|_\diamond$ are related in every specific case, together with Eq. (4) it would at least give a hint on how the sample size should be scaled with the number of involved qubits in order to achieve the desired uniformity of the distribution.

Conclusion and Outlook

Result 1 rigorously shows that we can efficiently simulate the Haar measure on compact Lie groups using even highly degenerate control systems as long as they satisfy the assumption on regularity. Our simulations in lower dimensions indicate that even with relatively large number of samples we quickly reach a point, after which we cannot distinguish the simulated distribution from the Haar measure in the second moments anymore. It would be interesting to see how this mixing time in the second moment scales in different systems with the sample size, or how fast we reach the mixing times in higher orders as compared to the second. For the precision with which we can hope to simulate the Haar measure in the k-th moment using N samples of any distribution, Eq. 4 suggests an order of $\mathcal{O}(\sqrt{d^k N^{-1}})$ in the 2-norm if we have $G \subset \mathrm{SU}(d)$. This is closely related to the question how to exactly quantify the desired precision, which still needs to be clarified. Nevertheless it seems to be a promising approach to be used for randomised benchmarking protocols such as in Table 1.

Part I

BACKGROUND ON QUANTUM INFORMATION AND QUANTUM CONTROL

In this part we provide the quantum-theoretic background needed to appreciate the results of Part III. The first Chapter gives a short introduction to quantum information. We proceed in Chapter 2 by briefly discussing control systems on compact Lie groups which are the tool to implement operations on quantum computers. Chapter 3 reviews the concept of unitary designs in digital quantum computing and formally generalise this to get the corresponding notion in the analogue setting.

1. FRAMEWORK OF QUANTUM INFORMATION

Quantum information differs form classical information procssing by the assumption that a binary piece of information is not necessarily in either of the classical states 0, 1 but rather in a quantum mechanical *superposition* of those. This has a significant impact on how to process information within such a framework. In particular one has to deal with certain obstacles such as the fact that quantum information cannot be copied. In this section, we provide a very short introduction to the basic framework of quantum information and its mathematical lanuage. As a general and comprehensive reference we hint at the book by Nielsen and Chuang [31].

Quantum States

In quantum mechanics, (pure) states are modelled as normalised elements of a complex Hilbert space $\mathcal{H}$. We will always consider the scalar product $\langle \cdot, \cdot \rangle$ to be antilinear in the *first* component. Depending on which property of a quantum system one is interested (e.g. spin of electrons, position of particle, polarisation of photons to name a few), the state space can be finite or infinite dimensional. Because we are mainly focusing on quantum information, all Hilbert spaces modeling quantum states are assumed to be *finite dimensional*. In particular we only deal with finite-dimensional representations of the involved Lie groups. Throughout we employ the Dirac notation where quantum states are denoted by 'kets' $|\psi\rangle \in \mathcal{H}$ and their canonical duals by 'bras' $\langle\psi| \in \mathcal{H}^*$. Scalar products are then expressed by $\langle \psi, \varphi \rangle = \langle \psi | \varphi \rangle$. If we choose an orthonormal basis $|1\rangle, \ldots, |d\rangle$, we can write a quantum state as $|\psi\rangle = \sum_i \alpha_i |i\rangle$. The intuition here is that if we perform a measurement in this basis, we find the system in state $|i\rangle$ with probability $|\alpha_i|^2$. These probabilities should add up to one of course, hence the condition on states being normalised is imposed.

Composite systems

For two quantum systems modelled by Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, the composite quantum system is given by the tensor product $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. A state $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ is called *seperable* if it can be decomposed into a tensor product $|\psi\rangle_{AB} = |\psi_A\rangle \otimes |\psi\rangle_B$. Otherwise it is called *entangled*. The concept of entangled states is the key ingredient for many potential application of quantum information, in particular for devising protocols for secure quantum key distribution. Entanglement will not explicitly play a role in our further treatment, but any introduction on quantum information would be substantially incomplete without it.

Evolution

For a closed system, quantum mechanics postulates that the time evolution is described by a unitary evolution, i.e. there exists a time-dependent unitary operator $U(t)$ such that for any state $|\psi(t)\rangle$ we have $|\psi(t)\rangle = U(t)|\psi(0)\rangle$. The evolution in a closed quantum systems is governed by

$$i\hbar\frac{d}{dt}|\psi(t)\rangle = H(t)|\psi(t)\rangle, \tag{1.1}$$

for some time-dependent self-adjoint operator $H(t)$, called the *Hamiltonian* of the system. Here $\hbar$ is the reduced Planck constant, which we consequently absorb by the Hamiltonian in the following.

The evolution is unitary only if the system does not interact with its environment. This is rather different if we perform *measurements* and thereby interfere with the system.

Measurements

Formally, a positive operator valued measurement (POVM) of a quantum system consists of a set of operators $M_1, \dots, M_k$ which are designed to detect certain outcomes $1 \dots, k$. The probability to detect outcome j is then $p(j) = \langle\psi|M_j^\dagger M_j|\psi\rangle$. Because all probabilities should add up to one, we impose the completeness relation $\sum M_j^\dagger M_j = \mathbb{1}$. An important special case are *projective measurements*, where we assume the measurement operators to be projectors P_j onto the eigenstates of some hermitian operator $A = \sum_j \lambda_j P_j$. Here the eigenvalues of A are the possible outcomes of the measurement. The expectation for the outcome when measuring A on the state $|\psi\rangle$ is given by $\langle A\rangle = \langle\psi|A|\psi\rangle$. We also see that measurements cannot detect global complex phases $e^{i\varphi}$ of $|\psi\rangle$. Therefore it suffices to model the evolution of quantum systems on $\mathrm{SU}(d)$.

How do we describe the state properly after having performed a measurement? The quantum mechanical superposition collapses randomly to one of the states $M_j|\psi\rangle$ with the respective probability during the process. This is structurally quite different from the original superposition, since we know that the system *is* in one of the states $M_j|\psi\rangle$ after the measurement, but we a priori do not know in *which*. In order to properly grasp this mathematically, we need to slightly enhance the model.

Mixed states

A *mixed state* is an operator $\rho \in L(\mathcal{H})$, sometimes also called *density operator*, which is selfadjoint, positive[1] and satisfies $\mathrm{Tr}(\rho) = 1$. It is called *pure* if there exists a state $|\psi\rangle \in \mathcal{H}$ such that $\rho = |\psi\rangle\langle\psi|$. Because they are positive and selfadjoint, mixed states can always be written as $\rho = \sum_i \lambda_i|\psi_i\rangle\langle\psi_i|$ with $\sum_i \lambda_i = 1$ for some orthogonal states $|\psi_i\rangle$. It is not difficult to show that a mixed state is pure if and only if $\mathrm{Tr}(\rho^2) = 1$.

[1] Recall that a selfadjoint operator A on a Hilbert space is positive iff $\langle v, Av\rangle \geq 0$ for all $v \in \mathcal{H}$

Quantum Channels

What conditions do we need to impose on a linear map $\Lambda : L(\mathcal{H}) \to L(\mathcal{H})$ such that mixed states are mapped into quantum states again? A *quantum channel* or *quantum map* on a Hilbert space $\mathcal{H}$ is a linear map $\Lambda : L(\mathcal{H}) \to L(\mathcal{H})$ which is

1. *trace preserving*, i.e. $\mathrm{Tr}(\Lambda\rho) = \mathrm{Tr}(\rho)$ for all $\rho \in L(\mathcal{H})$, and
2. *completely positive*, i.e. $\Lambda \otimes \mathbb{1}_n : L(\mathcal{H} \otimes \mathbb{C}^n) \to L(\mathcal{H} \otimes \mathbb{C}^n)$ is positive for every $n \in \mathbb{N}$.

Demanding complete positivity rather than just positivity is linked to the fact, that it should be possible to trivially extend quantum channels when considering composite systems. The structure of quantum channels is clarified by the next result, due to Kraus.

Theorem 1.1 (see [31], Theorem 8.1)**.** For any quantum channel acting on $L(\mathcal{H})$, there exists a finite collection of operators $\{M_1, \ldots, M_N\} \subset L(\mathcal{H})$ satisfying the completeness relation $\sum_j M_j^\dagger M_j = \mathbb{1}$, such that

$$\Lambda(\rho) = \sum_{j=1}^{N} M_j^\dagger \rho M_j \qquad \forall \rho \in L(\mathcal{H}).$$

Conversely, any such collection of operators defines a quantum channel.

Therefore *every* quantum operation can be perceived as some kind of measurement. This includes the unitary evolution as the special case with $N = 1$. A notion frequently used in quantum information to measure the distance of two quantum channels, is the metric associated to the *diamond norm*.

Definition 1.2. For a qunum channel $\Lambda : L(\mathcal{H}) \to L(\mathcal{H})$ we define the diamond norm by

$$\|\Lambda\|_\diamond := \sup_{n \in \mathbb{N}} \|\Lambda \otimes \mathbb{1}_n\|_{op} \tag{1.2}$$

where we equip $L(\mathcal{H} \otimes \mathbb{C}^n)$ with the 2-norm as usual.

With $\rho(t) = U_t \rho U_t^\dagger$, the mixed-state analogon of Schrödinger's equation reads

$$\frac{d}{dt}\rho(t) = -i[H, \rho(t)],$$

which is sometimes also called the *von Neumann equation*. This is one of the reasons why the Adjoint/adjoint representations of Lie groups and their algebras often appear in quantum mechanics.

Liouville Space

When working with quantum channels, it is sometimes convenient to use the *Liouville space*, which is a useful way of expressing linear operators acting on linear maps. The identification

$(v \otimes \beta)(w) = \langle \bar{\beta}, w \rangle v$ induces an isomorphism $\mathcal{H} \otimes \mathcal{H} \cong \mathrm{End}(\mathcal{H})$. With this we can express operators which act on $\mathrm{End}(\mathcal{H})$ by left or right multiplication with a fixed $A \in L(\mathcal{H})$ by $A \otimes \mathbb{1}$ and $\mathbb{1} \otimes A^T$ respectively. In this spirit, we occasionally write $U \otimes \bar{U}$ for the Adjoint representation of unitaries.

Pauli and Clifford groups

For the manipulation of one-qubit states, we can use the three Pauli-matrices

$$\sigma_x = \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ \sigma_y = \sigma_0 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},\ \sigma_z = \sigma_3 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

which correspond to the operations bitflip (σ_x), phaseflip (σ_z) and their combination (σ_y). It is important to observe that $\frac{i}{2}\sigma_x, \frac{i}{2}\sigma_y, \frac{i}{2}\sigma_z$ are cyclic generators of the (real) Lie algebra $\mathfrak{su}_2$: $[\frac{i}{2}\sigma_j, \frac{i}{2}\sigma_k] = \frac{i}{2}\epsilon_{jkl}\sigma_l$. The Pauli matrices generate a discrete subgroup of $\mathrm{U}(2)$, called the *Pauli group*. On n qubits the Pauli group is analogously defined to be the subgroup of $\mathrm{U}(2^n)$ generated by $\sigma_x, \sigma_y, \sigma_z$ acting locally on every site:

$$P_n := \{e^{ik\pi/2}\sigma_{j_1} \otimes \ldots \otimes \sigma_{j_n} | k, j_1, \ldots, j_n \in \{0,1,2,3\}\}.$$

However, the Pauli group P_n only provides very limited access to quantum operations. It only maps the elements in the computational basis into each other (up to factors) and is therefore insufficient to create entanglement. This can be accomlished by the *Clifford group*, which is the *normaliser* of the Pauli group

$$C_n := \{U \in \mathrm{U}(2^n) | U P_n U^\dagger \subset P_n\} / \mathrm{U}(1).$$

The Clifford group has the useful property, that with any other gate in $\mathrm{SU}(2^n)$ which is not already contained in C_n, it forms a *universal gate set* (Thm. 5.6 in [32]).

Definition 1.3. A set of quantum gates $D \subset \mathrm{SU}(d)$ is called *universal*, if the set

$$\{U_1 \cdots U_k | U_1, \ldots, U_k \in D,\ k \in \mathbb{N}\} \subset \mathrm{SU}(d)$$

is dense in $\mathrm{SU}(d)$.

We conclude this introduction with the remark that even though the Clifford group can create highly entangled states, this does not result in a significant advantage over classical computers. The remarkable *Gottesman-Knill Theorem* (Thm. 10.7 in [31]) states that circuits involving state preperation, Clifford gates and measurements can be simulated efficiently on classical computers. It is not yet fully understood which fundamental principles allow for instance Shor's factorisation algorithm to provide a considerable speedup compared to known classical algorithms.

2. QUANTUM CONTROL SYSTEMS

Quantum control theory and more importantly quantum *optimal* control are two fields of research that play a significant role in the development of efficiently operating quantum processors. In order to simulate different unitary operations within a quantum algorithm, it is necessary to manipulate the time-dependent Hamiltonian $H(t)$ governing the evolution inside the quantum system. The Hamiltonian $H(t)$ usually consists of a term H_0 inherent to the underlying platform (e.g. superconducting qubits or trapped ions to name a few) called the *drift* of the system, and some Hamiltonians $H_1(t), \dots, H_n(t)$ which can be tuned in strength and are referred to as the *controls*, resulting in the unitary evolution[1]

$$\frac{d}{dt}U(t) = iU(t)H(t) = iU(t)\left(iH_0 + \sum_j H_j(t)\right). \tag{2.1}$$

One major challenge of quantum control is linked to the fact that the dimension of the Hilbert state space scales exponentially with increasing number of qubits. Therefore it is not feasible to have control terms for all directions in the Hilbert space seperately, and we will usually find situations where the number of available controls is comparatively small. In order to overcome these degeneracies, we fundamentally rely on the non-commutativity of the control and drift Hamiltonians. Of course, finding paths in such a setting and even more so optimal ones, is significantly more challenging than in a non-degenerate case. Problems and methods related to finding such optimal solutions are more widely addressed within the framework of optimal control theory [33]. The control theory of quantum dynamical system is just the special case of this more general setting where we only consider *left-invariant* control systems on *compact* Lie groups. We want to emphasize that although we are restricting ourselves to this particular setting, most of the following definitions and result can be rephrased in a more general terms. For a detailed exposure on invariant control systems on Lie groups consider for instance [34, 35] and [36, 37, 38] for the more general treatment. Also we point at Appendix A for a summary of the definitions and results on Lie algebras which are used in this chapter.

Definition 2.1. Formally a left-invariant control system on a Lie group G is given by a collection $\Sigma = (X|Y_1, \dots, Y_m)$ of left-invariant vector fields on G, together with a subset of locally bounded

[1] Technically this defines the evolution in inverse time. We choose this convention since the vector fields $UH_0, \dots$ are left-invarian and becauset the commutator of matrices corresponds to the commutator induced by left-invariant vector fields on the Lie group $\mathrm{GL}(\mathbb{C}^d)$, whereas we would get an additional factor -1 in the right-invariant case.

and integrable control functions $F \subset L^1_{loc}(\mathbb{R}_{\geq 0}, \mathbb{R})$. This setup allows us to navigate along paths $\gamma : [0, \infty) \to G$ which are subject to a differential equation

$$\dot{\gamma}(t) = X(\gamma(t)) + \sum_j c_j(t) Y_j(\gamma(t)), \quad \gamma(0) = 0 \tag{2.2}$$

for some choice of control functions $c_1, \ldots, c_m \in F$. For the obvious reasons X is called the *drift* and $Y_1 \ldots, Y_m$ the *controls*. Further we define the *controlled distribution* $\mathcal{Y} \subset TG$ to be the distribution[2] spanned by the control vector fields $Y_1, \ldots, Y_m$ and $Lie(\mathcal{Y})$ its closure under the commutator. We occasionally write $X + \mathcal{Y} \subset TG$ for the affine distribution generated by the control system. Lastly, we also define

$$\mathcal{A}_T(g) := \{h \in G | \text{ there is a path } \gamma \text{ with } \dot{\gamma} \in X + \mathcal{Y} \text{ and } \gamma(0) = g,\ \gamma(T) = h\},$$

which is called the *attainable set* of the control system at time $T > 0$ from the starting point $g \in G$.

Remark 2.2. The notation $\Sigma = (X|Y_1, \ldots, Y_m)$ does not have an intrinsic motivation, we just use it to emphasize the distinction between the drift and the controls. If the left-invariant vector fields of Σ are given by $X = x^l, Y_1 = y_1^l, \ldots, Y_k = y_k^l$ for some elements in the Lie algebra $x, y_1, \ldots, y_k \in \mathfrak{g}$ we also employ the notation $\Sigma = (x|y_1, \ldots, y_m)$. Naturally, one can also consider right-invariant control systems which effectively leads to the same theory. It is clear that for a left-invariant system on a Lie group, the attainable sets for different starting points are related by $\mathcal{A}_T(g) = g \cdot \mathcal{A}_T(1_G)$ for all $g \in G$, and we only consider $\mathcal{A}_t := \mathcal{A}_t(1_G)$ in the following. We will not impose any restrictions on the control functions $F \subset L^1_{loc}(\mathbb{R}_{\geq 0}, \mathbb{R})$, other then being locally bounded and integrable. For many applications there might be restrictions, such as only allowing 'bang-bang controls' (alternating between ± 1) or piecewise constant functions. Using approximations if necessary, the latter turns out to be not much of an actual restriction. Since the control functions are only assumed to be measurable, all appearing differential equations should be understood to hold in the weak sense, i.e. under integration.

For a control system of interest, it is important to know which subset of the Lie group we can hope to reach. In the next step, one can then analyse how to steer the system in order to reach a given point in optimal time, usually under some restrictions on the available control functions. Regarding the quality of control over a dynamical system, we introduce the following distinctions.

Definition 2.3. A control system Σ on a Lie group G is called

1. *controllable* if $\mathcal{A}_\infty := \bigcup_{t \geq 0} \mathcal{A}_t = G$,

[2] In geometry, a distribution on a manifold M is a choice of subspaces of T_pM for every $p \in M$, which is spanned by a collection of smooth vector fields.

2. *finite-time* controllable if $\mathcal{A}_{\leq T} := \bigcup_{t \leq T} \mathcal{A}_t = G$ for some $T > 0$,

3. *exactly controllably at time* T if $\mathcal{A}_T = G$,

4. *strongly controllable* if $\mathcal{A}_{\leq T} = G$ for all times $0 < T < \infty$.

Assumption. In order to get controllability at all, we obviously need the respective Lie group to be *connected*. Therefore we will only work with connected Lie groups in the following, without further indication.

Any compact Lie group can be embedded into some unitary group $\mathrm{U}(d)$. Because quantum measurements are generally unable to detect global phases of states, we can restrict ourselves to analysing subgroups of $\mathrm{SU}(d)$. In particular, we say that a control system acting on a Hilbert space $\mathbb{C}^d$ is *fully controllable* if $\mathcal{A}_\infty = \mathrm{SU}(d)$.

For an invariant control system $(x|y_1, \ldots, y_m)$ on a Lie group G, we can naturally identify three subalgebras of $\mathfrak{g} = Lie(G)$ which are of significance for our discussion.

Definition 2.4 (compare [34], section 3)**.** Let $x, y_1, \ldots, y_m \in \mathfrak{g}$ be the generators of a left-invariant control system on some Lie group G. Then we define

1. the *dynamical Lie algebra* $\mathfrak{g}_\Sigma := \langle x, y_1, \ldots, y_m \rangle_{Lie} \subset \mathfrak{g}$,

2. the *zero-time ideal* $\mathfrak{I}_0 := \langle y_i, [y_i, x], i = 1, \ldots, m \rangle_{Lie} \lhd \mathfrak{g}_\Sigma$ generated by the controls $y_1, \ldots, y_m$, and

3. the *control subalgebra* $\mathfrak{g}_0 := \langle y_1, \ldots, y_m \rangle_{Lie} \subset \mathfrak{g}_\Sigma$.

Further we define $G_\Sigma, G_0 \subset G$ to be the subgroups generated by the Lie algebras $\mathfrak{g}_\Sigma$ and $\mathfrak{g}_0$, respectively. Clearly, we have the inclusions $\mathfrak{g}_0 \subset \mathfrak{I}_0 \subset \mathfrak{g}_\Sigma \subset \mathfrak{g}$. Note that $\mathfrak{I}_0$ is indeed an ideal, since it contains all commutators of the controls $y_1, \ldots, y_m$ with the drift, and therefore $[y_i, \mathfrak{g}_\Sigma] \subset \mathfrak{g}_\Sigma$ for all $i = 1, \ldots, m$. By construction, the codimension of $\mathfrak{I}_0 \subset \mathfrak{g}_\Sigma$ is at most 1, since all (iterated) commutators of x with the other generators y_i are automatically contained in $\mathfrak{I}_0$ and therefore $\mathfrak{g}_\Sigma = \mathrm{span}_{\mathbb{R}}\{x, \mathfrak{I}_0\}$.

On compact Lie groups, invariant control systems have the great advantage, that questions related to controllability can be answered completely in terms of these subalgebras, as the next proposition shows.

Proposition 2.5. Let $\Sigma = (x|y_1, \ldots, y_k)$ be an invariant control system on a compact Lie group G and $H \subset G$ a subgroup with Lie algebra $\mathfrak{h}$.

1. The condition $\mathfrak{g}_\Sigma = \mathfrak{h}$ is necessary and sufficient for the system to be *controllable* on H in which case it is also finite-time controllable.

2. Under the stronger assumption $\mathfrak{g}_0 = \mathfrak{h}$, the system is also *strongly controllable* on H.

We will primarily comprehend (strong) controllability in terms of these algebraic conditions. The condition $\mathfrak{g}_0 = \mathfrak{h}$ in the second claim should also be necessary, despite some effort on my own and an extensive search in the literature, I could not find a simple proof for this conjecture.

Proof. The first statement is a standard result - known as Chow's Theorem - and can be found for instance in Theorem 7.1 of [34].
Claim 2. should also be a standard result, but since I could not find a formal proof in the literature we proceed to state it here. The second claim clearly holds if the system is *homogenous*, i.e. we have $x = 0$. By rescaling the control functions if necessary, we can then always find a path connecting two points in arbitrarily small time $T > 0$. Now consider the case where a drift is present. Fix a time $T > 0$ and a point $p \in H$ that we want to reach from 1_H. Let $c_1, \dots, c_k$ denote a choice of controls, which steer the *homogenous* system in time T from 1_H to p over a path γ. Consider the sequence $p_n := \gamma_n(\frac{T}{n})$ with $n \in \mathbb{N}$, where γ_n is the unique solution to $\dot{\gamma}_n(t) = X(\gamma_n(t)) + \sum_j n\, c_j(nt) Y_j(\gamma_n(t))$ with $\gamma_n(0) = 1_H$. In order to show that $p_n \to p$, consider H to be given in some faithful representation, such that we can exploit the surrounding euclidean structure. Let $C_n(t) \geq 0$ denote the Lipshitz constant of the time-dependent vector field $\sum_j n c_j(tn) Y_j$ for $0 \leq t \leq T/n$, and without loss of generality assume that $\operatorname{diam}(H), \|X\| \leq 1$. We can then estimate

$$\begin{aligned} \frac{d}{dt}\|\gamma(nt) - \gamma_n(t)\|^2 &= \langle n\dot{\gamma}(nt) - \dot{\gamma}_n(t), \gamma(nt) - \gamma_n(t)\rangle \\ &\leq C_n(t)\|\gamma(nt) - \gamma_n(t)\|^2 + \|X\| \operatorname{diam}(H). \end{aligned} \tag{2.3}$$

Now we use an argument similar to the proof of Grönwall's Lemma [39]. Consider the functions $v(t) := e^{\int_0^t C_n(s)ds}$ and $u(t) := \|\gamma(nt) - \gamma_n(t)\|^2$. With (2.3) we can estimate

$$\frac{d}{dt}\left(\frac{u(t)}{v(t)} - t\right) = \frac{u'(t) - C_n(t)u(t)}{v(t)} - 1 \leq 0$$

for all $t \leq T/n$. Therefore the function $\frac{u(t)}{v(t)} - t$ is decreasing in t and we have

$$\frac{u(T/n)}{v(T/n)} - \frac{T}{n} \leq u(0)$$

which is equivalent to

$$u\left(\frac{T}{n}\right) \leq \left(u(0) + \frac{T}{n}\right) e^{\int_0^{T/n} C_n(s)ds}.$$

Since $\int_0^{T/n} C_n(s)ds = \int_0^T C_1(s)ds$ is independent of n and we have $u(0) = 0$ this shows that indeed $p_n \to p$. In order to conclude the proof, we use the fact that the attainable set $\mathcal{A}_t \subset H$ has nonempty interior for all $t > 0$ if $\mathfrak{g}_0 = \mathfrak{h}$ holds (cf. Thm. 3.2 in [38]). □

It is self-evident, that strong controllability is a significantly more refined property than just hav-

ing controllability of the system. One major structural difference is the following: The controlled distribution $\mathcal{Y}$ of a strongly controllable system $\Sigma = (X|Y_1, \ldots, Y_m)$ satisfies $Lie(\mathcal{Y})_g = T_g G_\Sigma$, which is the defining property of so called *sub-Riemannian structures*. In this case the control distance on G_Σ associated on by

$$d_\Sigma(p,q) := \inf \{\text{Length of all curves } \gamma \text{ which start at } p, \text{ end in } q, \text{ and satisfy } \dot{\gamma} \in X + \mathcal{Y}\}.$$

is a proper metric, since the influence of the drift can be neglected following the arguments in the proof of Prop. 2.5. In general d_Σ is not symmetric, because we need to compensate for the drift when reversing the direction. For many applications however, strong controllability - in the sense that $\mathfrak{g}_0 = \mathfrak{g}_\Sigma$ holds - is not what we would usually encounter. As far as quantum control setups are concerned, the control Hamiltonians are usually more local in nature, such as single-qubit operations or nearest-neighbour interactions. The drift on the other hand is used to couple these subsystems and therefore plays a decisive role in generating the dynamical Lie algebra.

The next proposition suggests that the 'intermediate case' in which $\mathfrak{I}_0 = \mathfrak{g}_\Sigma$ is satisfied can serve as a substitute for strong controllability, at least to some extent. Specifically, we still get some results related to exact controllability.

Proposition 2.6. Consider a left-invariant control system $\Sigma = (X|Y_1, \ldots, Y_k)$ on a Lie group G. Then the following are eqivalent:

1. The zero time-ideal is equal to the generated Lie algebra: $\mathfrak{I}_0 = \mathfrak{g}_\Sigma$.
2. Σ has the *strong accessibility property*, that is $\mathcal{A}_T$ has nonempty interior in G_Σ for every $T > 0$.

If G_Σ is compact, 1. and 2. are also equivalent to:

3. There exists a time $T > 0$ such that $\mathcal{A}_t = G_\Sigma$ holds for all $t \geq T$.

Proof. The equivalence $1. \Leftrightarrow 2.$ is proved in more general terms in [38] (Thm. 3.2) and $1. \Leftrightarrow 3.$ can be found for instance in [40] (Thm. 4). □

Remark 2.7. At the core, the Propositions 2.5 and 2.6 do not actually require the setting of Lie groups. All these assertions can be rephrased in more general terms on arbitrary manifolds using the Lie subalgebras generated by the vector fields which are defined accordingly. The equalities are then replaced by a spanning property, in the sense that the respective Lie algebras of vector fields pointwise span the tangent space of the manifold considered. Only the converse implications $3. \implies 1., 2.$ of Prop. 2.6 are generally lost in this case, see [40].

With the second characterisation of Proposition 2.6, we see that the condition $\mathfrak{I}_0 = \mathfrak{g}_\Sigma$ essentially informs us if we can influence the drift direction by varying the controlled parts.

Consider for instance a control system $\Sigma = (x|y_1, \ldots, y_m)$ generating a compact Lie group G such that the center $\mathfrak{z}(\mathfrak{g}_\Sigma)$ is non-trivial, satisfies $\mathfrak{z}(\mathfrak{g}_\Sigma) \cap \mathfrak{g}_0 = \{0\}$ and that $x \in \mathfrak{z}(\mathfrak{g}_\Sigma)$. Then the attinable set at time $t > 0$ is contained in $\mathcal{A}_t \subset \{\exp(tx)\} \times_\Sigma G/\mathcal{Z}(G_\Sigma)$ which is a submanifold of G_Σ with codimension greater or equal to one [3]. Therefore we give the condition $\mathfrak{I}_0 = \mathfrak{g}_\Sigma$ a proper name.

Definition 2.8. A left-invariant control system $\Sigma = (x|y_1, \ldots, y_m)$ on a Lie group G is called *regular*, if it satisfies $\mathfrak{I}_0 = \mathfrak{g}_\Sigma$. Accordingly, we call the corresponding collection of generators $(x|y_1, \ldots, y_m)$ a *regular generating set*[4] for $\mathfrak{g}_\Sigma$ in this case.

Remark 2.9. An invariant control system $(x|y_1, \ldots, y_m)$ on $\mathfrak{g}$ automatically satisfies the regularity assumption if the generated algebra $\mathfrak{g}_\Sigma$ is perfect, that is we have $[\mathfrak{g}_\Sigma, \mathfrak{g}_\Sigma] = \mathfrak{g}_\Sigma$. This is because the brackets between the generators satisfy $[x, x], [x, y_i], [y_i, y_j] \in \mathfrak{I}_0$ for all $i, j = 1, \ldots, n$ we can then conclude $\mathfrak{g}_\Sigma = [\mathfrak{g}_\Sigma, \mathfrak{g}_\Sigma] \subset \mathfrak{I}_0$. In particular this holds if $\mathfrak{g}_\Sigma$ is *semisimple*. At this point it is worthwile to recall that a real compact Lie algebra is semisimple if and only if it is perfect, see Prop. A.30.

An example of generators which do not satisfy our regularity condition is the following.

Example 2.10. Consider the Lie algebra which is obtained by trivially extending $\mathfrak{su}_2$:

$$\mathfrak{g} := \mathfrak{su}_2 \oplus \mathbb{R} \tag{2.4}$$

with $[x + \lambda, y + \mu] = [x, y]$ for all $x, y \in \mathfrak{su}_2$ and $\lambda, \mu \in \mathbb{R}$. Let x, y, z be cyclic generators of $\mathfrak{su}_2$. We observe the following:

1. *There control system $\Sigma = (z + \lambda|x)$ is non-regular.*
   For any $\lambda \neq 0$ we have $\mathfrak{g}_\Sigma = \langle x, z + \lambda \rangle_{Lie} = \mathfrak{g}$ but $\mathfrak{I}_0 = [\mathfrak{g}_\Sigma, x] = \mathfrak{su}_2 \neq \mathfrak{g}_\Sigma$.

2. *The definition depends decisively on the order of the generators.*
   Upon exchanging the roles of x and $z + \lambda$, we do get $\mathfrak{I}_0 = [z + \lambda, \mathfrak{g}_\Sigma] = \mathfrak{g}_\Sigma$. Obviously $[\mathfrak{g}_\Sigma, \mathfrak{g}_\Sigma] = \mathfrak{su}_2 \neq \mathfrak{g}_\Sigma$ which shows that for regularity it is not necessary that $\mathfrak{g}_\Sigma$ is perfect.

Example 2.11. A concrete representation of Example 2.10 is for instance given on a two-qubit system by

$$\phi : \mathfrak{su}_2 \oplus \mathbb{R} \to \mathrm{End}(\mathbb{C}^2)^{\otimes 2}, \qquad \phi(x + \lambda) = x \otimes \mathbb{1}_2 + i\lambda \mathbb{1}_2 \otimes \sigma_z.$$

This corresponds to writing the Lie algebra (2.4) more suggestively as $\mathfrak{g} = \mathfrak{su}_2 \oplus i\mathbb{R}$ since the trivial part exponentiates to a compact Lie group under this representation.

[3] In fact it is exactly one, but that is not important here.

[4] Technically, this is *not* a set, but rather a pointed set, that is a set with a special element which should be treated differently than the others.

Finally, the Lie algebra (2.4) also appears in an example from physics - specifically the two-qubit Hubbard model we used in Example of Fig. 4-D in our simulations.

Example 2.12. Consider the two-qubit Hubbard model with an Heisenberg XY-interaction and transverse magnetic field, that is the drift of the control system is given by

$$H_0 = \sigma_x^1\sigma_x^2 + \sigma_y^1\sigma_y^2 + \sigma_z^1 + \sigma_z^2,$$

and we choose $H_c := \sigma_x^1\sigma_x^2$ as single control. Let $\mathfrak{g}_\Sigma := \langle iH_0, iH_c\rangle_{Lie}$ be the dynamical Lie algebra. If we substitute $h_+ := -\frac{i}{2}\sigma_x^1\sigma_x^2$, $h_- := -\frac{i}{2}\sigma_y^1\sigma_y^2$, $e := \frac{i}{2}\left(\sigma_x^1\sigma_y^2 + \sigma_y^1\sigma_x^2\right)$ and $f := \frac{i}{2}\left(\sigma_z^1 + \sigma_z^2\right)$, we get the commutation relations

$$[h_\pm, e] = \pm f, \quad [h_\pm, f] = \mp e, \quad [h_+, h_-] = 0, \quad [e, f] = 2(h_+ - h_-). \tag{2.5}$$

In particular $iH_0 = -2(h_+ + h_- - f)$, $iH_c = -2h_+$, therefore we have

$$\frac{1}{4}[iH_c, iH_0] = -[h_+, h_+ + h_- - f] = [h_+, f] = -e, \quad [iH_c, e] = -2f,$$

and get $\langle iH_c, iH_0\rangle_{Lie} = \mathrm{span}_{\mathbb{R}}\{h_\pm, e, f\}$. It is easy to see that the elements $h_+ - h_-, e, f$ generate a Lie algebra isomorphic to $\mathfrak{su}_2$ and that $h_+ + h_- \in \mathfrak{z}(\mathfrak{g}_\Sigma)$. Therefore the dynamical Lie algebra $\mathfrak{g}_\Sigma$ is indeed isomorphic to $\mathfrak{su}_2 \oplus \mathbb{R}$. For our choice of generators, the regularity condition $\mathfrak{I}_0 = \mathfrak{I}(h_+) = \mathfrak{g}_\Sigma$ is satisfied, which is easily seen from the commutation relations (2.5).

How can we determine if we can control a subgroup $H \subset \mathrm{SU}(d)$ with a control system $\Sigma = (x|y_1 \ldots, y_m)$ on $\mathrm{GL}(\mathbb{C}^d)$? Determining the generated Lie subalgebra using computer algebra systems becomes quite expensive pretty fast. One possible approach here is to analyse the joint *symmetries* of the generators, i.e.

$$\mathfrak{com}(\Sigma) = \mathfrak{com}(x, y_1, \ldots, y_m) = \{z \in \mathrm{End}(\mathbb{C}^d) | [z, x_0] = [z, y_i] = 0 \text{ for all } i\},$$

which can be handled a bit easier computationally. In order to have controllability on H it is necessary that the spaces describing these symmetries coincide, i.e. $\mathfrak{com}(\Sigma) = \mathfrak{com}(\mathfrak{h}) \subset \mathrm{End}(\mathbb{C}^d)$. It is not difficult to accept that this is not sufficient in general. Consider for instance the real orthogonal matrices $\mathrm{SO}(d) \subset \mathrm{SU}(d)$ inside the unitaries. For both groups the only endomorphisms that commute with all elements are multiples of the identity $\mathbb{1}_d$, but clearly $\mathrm{SO}(d)$ is a proper subgroup of $\mathrm{SU}(d)$. In order to get a suffiicent criterion for controllability, we need to take the symmetries of the *tensor square* of the standard representation into account as well.

Theorem 2.13 (part of Thm. 15 in [26])**.** Let $\mathfrak{h}$ be a subalgebra of a compact semisimple Lie algebra $\mathfrak{g}$ and consider a faithful representation ϕ of $\mathfrak{g}$. Then the following statements are equivalent:

1. $\mathfrak{h} = \mathfrak{g}$,
2. $\mathfrak{com}_{\phi\otimes\phi}(\mathfrak{h}) = \mathfrak{com}_{\phi\otimes\phi}(\mathfrak{g})$.

For the complete results as well as more details on this, we refer to [41, 26].

Remark 2.14. In the following chapters, we will always drop the subscript Σ, by restricting the surrounding Lie group if necessary. Specifically, if we have e.g. $x, y_1, \ldots, y_m \in \mathfrak{su}_d$ which generate a Lie subalgebra $\mathfrak{g} \subset \mathfrak{su}_d$, we assume that the corresponding control system lives only on $\mathfrak{g}$ and the generated Lie subgroup $G \subset \mathrm{SU}(d)$ right away.

3. UNITARY DESIGNS AND K-HAAR MEASURES

Throughout the following, we are concerned with a compact, connected (non-abelian) Lie group G. For a probability measure ν on G and a finite-dimensional representation $\Phi : G \to \mathrm{GL}(V)$, we define the moment of ν with respect to Φ by

$$M_\nu^\Phi := \int_G \Phi_g d\nu(g). \tag{3.1}$$

The respective moment of the Haar measure μ_G will mostly be denoted as M_G^Φ. For the moments M_ν^Φ taken with respect to tensor powers of the standard Adjoint representation $\Phi = \Phi^{k,k}$ (see Rem. A.23) we simply write M_ν^k. Recall that we can turn any finite-dimensional representation of $\Phi : G \to \mathrm{GL}(V)$ into a unitary representation via

$$\langle x, y\rangle_\Phi := \int_G \langle \Phi_g x, \Phi_g y\rangle d\mu_G(g),$$

starting with any scalar product $\langle \cdot, \cdot \rangle$ on V. In particular we get

$$\|M_\nu^\Phi\|_{op} \leq \sup_{0\neq v\in\mathbb{C}^d} \frac{1}{\|v\|} \int \|\Phi_g v\| d\nu \leq 1. \tag{3.2}$$

3.1 Unitary Designs

In mathematics the theory of *combinatorial designs* is dedicated to the study of finite sets that satisfiy some requirements of symmetry and/or balance in various structures. For instance by considering a regular polyeder with edges on the unit sphere in $\mathbb{R}^3$, an integration of polynomials up to a certain degree over the sphere can be realised by averaging the polynomial over those edges [42]. This is an example for structures which are known as *spherical designs*. In the context of quantum information, *unitary designs* were introduced in a thesis by C. Dankert [43] who transferred the notion of such spherical designs to the setting of unitary groups. Formally a unitary k-design[1] is a *finite* set $D = \{u_1 \ldots, u_n\} \subset \mathrm{SU}(d)$ such that the $\Phi^{k,k}$-th moments of the uniform

[1] In the literature, these are usually referred to as *t-designs*. To avoid any confusions with the time paramter t we shall stick to the term k-design throughout the remainder of this thesis.

distribution over D and the Haar measure μ_G coincide:

$$M_D^k = \frac{1}{|D|^2} \sum_{U \in D} U^{\otimes k} \otimes \bar{U}^{\otimes k} = \int U^{\otimes k} \otimes \bar{U}^{\otimes k} d\mu_G(U) = M_G^k. \tag{3.3}$$

Surprisingly such unitary designs exist for all Several discrete gate sets were subsequently identified a unitary 2 and 3-designs, most prominently the Clifford group [16]. Beyond this however, the involved discrete group structures become increasingly complicated. For many applications it completely suffices to relax this definition by demanding the equality (3.3) only to hold up to a small error ε in a norm of choice, leading to the notion of *approximate* designs. Approximate designs of arbitrary error and order $k \in \mathbb{N}$ can be for instance generated by considering finite products of elements in a universal gate set (Def. 1.3) up to a certain length.

The implementation of quantum gates is generally not trivial and it seems unnecessarily complicated to exactly implement gates which are supposed to simulate a random distribution in the first place. Since on a quantum computer discrete gate sets are generated analogously using control systems, the idea to consider continuous analogues of unitary designs comes quite natural. This approach has already been investigated in several publications, e.g. [18, 19, 44, 13]. Since the defining property of designs is to be *finite* subsets simulating the uniform distribution up to some moment, naming such continuous distributions 'designs' is somewhat abusive. Therefore we introduce a different naming and investigate which equivalent characterisations of unitary designs (e.g. [29]) carry over to the continuous setting analogously.

3.2 k-Haar Measures

Definition 3.1. Let G be a compact (non-abelian) Lie group with a representation $\Phi : G \to \mathrm{GL}(V)$ on a Hilbert space V, and μ_G the respective normalised Haar measure. A measure ν on G is called a Φ-*Haar measure* if the respective moment coincides with the one of the Haar measure, i.e.

$$M_\nu^\Phi = M_G^\Phi. \tag{3.4}$$

If $\Phi = \Phi^{k,k}$, we simply say that ν is a k-Haar measure.

Again, we can relax this definition by considering Eq. (3.4) only to hold approximately in a suitably chosen norm. We will mostly consider the operator norm, which is in some sense immune to scaling of the dimensions both of G and the representation (3.2), and the 2-norm which is easier to handle computationally. However there is some ambiguity in the choice here, as we already discussed in the introduction.

Definition 3.2. A measure ν on G as above is called an ε-approximate Φ-Haar measure if $||M_\nu^\Phi - M_G^\Phi||_{op} \leq \varepsilon$.

In any case, it is useful to determine the moments of the Haar measure in the first place.

Proposition 3.3. The moment M_G^Φ of the unitary representation $\Phi : G \to \mathrm{GL}(V)$ is the orthogonal projection from V onto the subspace V^G that invariant under the G-action.

Proof. Because $M_G^\Phi \circ M_G^\Phi = M_{\mu_G * \mu_G}^\Phi = M_G^\Phi$ it is clear that M_G^Φ is a projection. Since $M_G^\Phi v = \int_G \Phi_g v d\mu_G = v$ for all $v \in V^G$ we have that the image of M_G^Φ contains V^G. Conversely,

$$\Phi_g \circ M_G^\Phi v = \int \Phi_g \Phi_h v d\mu_G(h) = \int \Phi_{gh} v d\mu_G(h) = \int \Phi_h v d\mu_G(h) = M_G^\Phi v. \tag{3.5}$$

shows that V^G contains the image of M_G^Φ and therefore M_G^Φ is indeed a projection onto the invariant subspace V^G. For the orthogonality, it suffices to check that M_G^Φ is self-adjoint. Using unimodularity of the Haar measure we compute

$$\left(M_G^\Phi\right)^\dagger = \int \Phi_g^\dagger d\mu_G(g) = \int \Phi_{g^{-1}} d\mu_G(g) = \int \Phi_g d\mu_G(g) = M_G^\Phi,$$

which concludes the proof. □

The following theorem is inspired by various equivalent chracterisations of unitary 2-designs from [29] (compare Def. 1, Prop. 1, Thm. 3 therein).

Theorem 3.4. Let $\Phi : G \to \mathrm{GL}(V)$ be a representation of a compact Lie group and μ_G the Haar measure. Then for any probability measure ν on G, the following are equivalent:

1. $M_\nu^\Phi = M_G^\Phi$,
2. The 2-norms of M_ν^Φ and M_G^Φ coincide, which is by Prop. 3.3, the same as

$$\|M_\nu^\Phi\|_2 = \left(\dim(V^G)\right)^{1/2}. \tag{3.6}$$

For $\Phi = \Phi^{k,k}$, these are also equivalent to

3. For every polynomial p of homogenous degree (k,k) in the coordinates of $U \in \mathrm{SU}(d)$ and their complex conjugates, $p(U) = p(U_{ij}, \bar{U}_{ij})$, we have

$$\int_G p(U_g) d\nu = \int_G p(U_g) d\mu_G. \tag{3.7}$$

Theorem 2 in [29] is really just the equivalence 1. $\Leftrightarrow$ 2. in disguise by computing the squared 2-norms respectively.

Proof. $(1) \Leftrightarrow (2)$

$$\begin{aligned}\|M_\nu^\Phi - M_G^\Phi\|_2^2 &= \mathrm{Tr}((M_\nu^\Phi - M_G^\Phi)\left(M_\nu^\Phi - M_G^\Phi\right)^\dagger) \\ &= \mathrm{Tr}\left((M_\nu^\Phi)^\dagger M_\nu^\Phi - M_\nu^\Phi \left(M_G^\Phi\right)^\dagger - M_G^\Phi \left(M_\nu^\Phi\right)^\dagger + \left(M_G^\Phi\right)^\dagger M_G^\Phi\right).\end{aligned}$$

With the understanding that M_G^Φ is self-adjoint (it is an orthogonal projection) and that $M_{\mu_G}^\Phi M_\nu^\Phi = M_{\mu_G * \nu}^\Phi = M_{\mu_G}^\Phi$, this reduces to

$$\|M_\nu^\Phi - M_G^\Phi\|_2^2 = \|M_\nu^\Phi\|_2^2 - \|M_G^\Phi\|_2^2. \tag{3.8}$$

Clearly, if (1) holds, then the norms on the right hand side of (3.8) coincide, and vice versa.
$(1) \Leftrightarrow (3)$: The Kronecker product of two matrices A, B is given by

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & & \vdots \\ a_{n1}B & \dots & a_{nn}B \end{pmatrix}.$$

Therefore every (k,k)-homogenous polynomial in the entries of U can be found as a linear combination of the entries of $U^{\otimes k} \otimes \bar{U}^{\otimes k}$, which proves the last equivalence. □

It really is the last statement of Thm. 3.4 that makes k-Haar measures useful for computations. In particular 2-Haar measures on $\mathrm{SU}(d)$ already prove to be helpful in many cases, since we have exact integration rules for $(2,2)$-homogenous polynomials (Prop. 4.2.3 in [45]). The next result shows that we can in some sense reduce the 'order' of a Φ-Haar measure.

Proposition 3.5. Let $\Phi : G \to \mathrm{GL}(V)$, $\Psi : G \to \mathrm{GL}(W)$ be two representations of a compact Lie group G. If ν is a $\Phi \otimes \Psi$-Haar measure on G and the invariant subspace W^G is non-trivial, then ν is also a Φ-Haar measure.

Proof. Take a non-zero $w \in W^G$ and $v \in V$. Because we have $\Psi_g(w) = w$ for all $g \in G$ we get

$$M_\nu^{\Phi\otimes\Psi}(v \otimes w) = \int_G \Phi_g(v) \otimes \Psi_g(w)\, d\nu(g) = \left(\int_G \Phi_g(v) d\nu(g)\right) \otimes w = M_\nu^\Phi(v) \otimes w.$$

Clearly, $M_\nu^{\Phi\otimes\Psi} = M_G^{\Phi\otimes\Psi}$ then implies $M_\nu^\Phi = M_G^\Phi$, which concludes the proof. □

Corollary 3.6. In the special case where we consider the representation $\Phi = \Phi^{k,k}$ on $V = \mathrm{End}(\mathbb{C}^d)^{\otimes k}$, we have $\mathbb{1}_{d^k} \in VG$. Thus a k-Haar measure on G is also a $(k-1)$-Haar measure and inductively also for all $1 \leq m \leq k$.

Remark 3.7. For our simulations in the introduction we rely on characterisation 2. of Thm. 3.4. In order to avoid memory issues when computing the 2-norm of the moments of the simulated

distributions, we rely on the following computation

$$\begin{aligned}\|M_\nu^k\|_2^2 &= \operatorname{Tr}\left(\left(\int U_1^{\otimes k}\otimes \bar{U}_1^{\otimes k} d\nu(U_1)\right)\left(\int U_2^{\dagger\otimes k}\otimes \bar{U}_2^{\dagger\otimes k} d\nu(U_2)\right)\right)\\ &= \int\int \operatorname{Tr}\left((U_1U_2^\dagger)^{\otimes k}\otimes(\overline{U_1U_2^\dagger})^{\otimes k}\right) d\nu(U_1)d\nu(U_2)\\ &= \int\int |\operatorname{Tr}(U_1U_2^\dagger)|^{2k} d\nu(U_1)d\nu(U_2). \end{aligned} \tag{3.9}$$

The expression on the right is the continuous analogue of the *frame potential*, as defined in [29]. Therefore we can determine the 2-norm error of the k-th moment by comparing the frame potential of the distribution with the number of symmetries $\dim(V^G)$ of the corresponding Lie algebra representation $\Phi^{k,k}$.

Part II

STOCHASTIC ANALYSIS ON LIE GROUPS

In the context of quantum control, we are interested in studying systems that are exposed to random control profiles. By considering idealised models, the resulting differential equations are driven by Brownian motion and require special treatment. *Chapter 4* introduces the two most important notions of stochastic integration and stochastic differential equations (SDEs), which are due to Itô and Stratonovich. We continue by discussing their properties, as well as their respective advantages and inherent subtleties. Depending on the situation, generally one of the two is preferrable over the other. However, we can relate Itô and Stratonovich integration in a large class of cases, allowing us to exploit the advantages of both. In particular this is possible when dealing with stochastic differential equations.

Chapter 5 continues by discussing properties of processes that arise as solutions to left-invariant and time homogenous stochastic differential equations on Lie groups. We a particularly focus on Markov properties, associated probability semigroups and generators which allow to determine the long-term behaviour of such processes.

4. STOCHASTIC DIFFERENTIAL CALCULUS

We use the notations and conventions introduced in Appendix B. In particular, $(\Omega, \mathcal{P}, (\mathscr{F}_t)_{t\geq 0})$ always denotes the underlying filtered probability space on which we model all stochastic processes. We assume basic familiarity with stochastic processes, Brownian motions, and martingales or point at Appendix B for a short summary. Itô's construction prominently involves the *quadratic variation process* of square integrable martingales (Def. B.28). Recall that the quadratic variation of a square-integrable martingale $(M_t)_t$ is the unique increasing, right-continuous process $([M]_t)_t$ such that $M^2 - [M]$ is again a martingale. For a Brownian motion $(B_t)_t$ it is simply given by $[B]_t = t$, see the remark after Def. B.28. Because the quadratic variation is increasing and right-continuous its sample paths $[M](\omega)$ are the cumulative distribution function of a measure $d[M](\omega)$ on $\mathbb{R}_{\geq 0}$, where we omit the dependence on $\omega \in \Omega$ in the following. Similarly, the cross variation[1] $[M, N] = ([M + N] - [M - N])/4$ of two martingales M, N is the cumulative distribution function of a *signed* measure $d[M, N]$, where we allow negative weights. Note that the cross variation is zero if the processes are independent. All appearing martingales are supposed to be square integrable in the following. For a comprehensive overview of stochastic integration and SDE theory we point out [46, 47].

4.1 Stochastic Integration

In order to infinitesimally characterise stochastic processes which are driven by some noise process $(X_t)_t$, we first try to find a way to integrate functions with respect to sample paths of X. The naive approach here would be to take a sample path $X_t(\omega)$ and a function $f : \mathbb{R}_{\geq 0} \to \mathbb{R}$ (which also depends on ω) and proceed in the spirit of the Riemann-Stieltjes construction. We consider a sequence of approximating sums for every $\omega \in \Omega$ seperately and define the integral as the limit

$$\int_0^T f_s(\omega) dX_s(\omega) := \lim_{\|\Pi\|\to 0} \sum_{j=1}^{n} f_{s_k}(\omega)(X_{t_{j+1}}(\omega) - X_{t_j}(\omega)), \tag{4.1}$$

where Π are partitions $\{t_1, \ldots, t_n\}$ of the interval $[0, T]$ with $\|\Pi\| = \max_j |t_{j+1} - t_j|$ and $s_k \in [t_j, t_{j+1})$. In order for this limit to be well defined in the sense that it converges to the same value for all choices $s_k \in [t_j, t_{j+1})$, it is in general necessary that the function f is locally of

[1] The notation $[\cdot, \cdot]$ is also employed for the Lie bracket in the later chapters. Whether $[\cdot, \cdot]$ should be understood in this sense or as the cross-variation is clear from the context in every case.

bounded variation. This means that the limit

$$\lim_{N\to\infty} \sup_{0<t_1\cdots<t_n<t} \sum_{j=1}^{N} |f(t_{j+1}) - f(t_j)| < \infty \tag{4.2}$$

is finite for all times $t \geq 0$. Therefore, if the sample paths of the process X have bounded first variation almost surely, this approach works perfectly fine. Unfortunately for many interesting processes such as martingales and in particular Brownian motion, the first variation is infinite almost surely and therefore, the limit (4.1) is not well-defined in this sense.

In order to circumvent this obstacle, we weaken the notion of convergence, by always evaluating the integrand at one specific point of the intervals in the approximation. As first discovered and rigorously proven by Kyoshi Itô [48], with a careful choice of this point, one can exploit the fact that the quadratic variation of square-integrable martingales is finite. This is sufficient to show that the limit (4.1) then exists almost surely for sample paths of a Brownian motion, as long as the integrand is well behaved enough. However, the construction now significantly depends on the specific point of the intervals where we choose to evaluate the integrand.

From this ambiguity, two important notions of stochastic integration arise: the *Itô* integral which evaluates the integrand at the *lower end* of the interval, and on the other hand the *Stratonovich* integral - originally introduced by the Russian physicist Ruslan Stratonovich [49] - which uses the value in the *middle*. Both have their advantages and subtleties which we are going to discuss in the following. A detailed construction of the stochastic Itô integral can be found for instance in [46] or in Chapter 5 of [47]. Chapter 8 of [47] also features a nice exposition of Stratonovichs construction.

Both approaches provide methods to perform stochastic integration with respect to square-integrable martingales. For procceses which have bounded first variation, the Riemann-Stieltjes construction works as usual. Therefore the most general type of stochastic processes $(X_t)_t$ for which we can make sense of stochastic integration are of the type

$$X_t = X_0 + V_t + M_t \tag{4.3}$$

where $(V_t)_t$is an adapted process with bounded first variation and $(M_t)_t$ a local square-intagrable martingale, such that $M_0 = V_0 = 0$. Such processes are known as *semimartingales*. Similar to the construction of the Riemann integral, we begin by looking at integrands which have piecewise constant sample paths.

Definition 4.1. A stochastic process $(A_t)_t$ is called *simple* if there is an increasing sequence of times $0 = t_0 < t_1 < ...$, such that it can be written as a sum of characteristic functions

$\chi_{[t_k,t_{k+1})} : [0,\infty) \to \{0,1\}$ with random coefficients $a_k : \Omega \to \mathbb{R}$, i.e.

$$A_s(\omega) = \sum_{k=0}^{\infty} a_k(\omega)\chi_{[t_k,t_{k+1})}(s). \tag{4.4}$$

Definition 4.2 (Itô integral for simple integrands)**.** Let $(M_t)_t$ be a martingale, and $(A_t)_t$ a simple process as in (4.4) adapted to M. For $t \geq 0$ we define the Itô integral to be the random variable

$$\int_0^t A_s dM_s := \left(\sum_{k=0}^{N_t-1} a_k(M_{t_{k+1}} - M_{t_k}) \right) + a_{N_t}(M_t - M_{t_{N_t}}) \tag{4.5}$$

with $N_t = \max\{k : t_k \leq t\}$.

Definition 4.3 (Stratonovich integral for simple integrands)**.** In the same situation, we define the Stratonovich integral for a simple process by

$$\int_0^t A_s \odot dM_s := \left(\sum_{k=0}^{N_t-1} \frac{a_{k+1}+a_k}{2}(M_{t_{k+1}} - M_{t_k}) \right) + \frac{a_t + a_{t_{N_t}}}{2}(M_t - M_{t_{N_t}}).$$

Remark 4.4. The notation $A \odot dM$ has no particular intrinsic meaning, we just use it to make a clear distinction between the two different constructions. In the integrals, the subscripts are supposed to indicate that the parameter we are integrating against is essentially the time and the process just alters the weight of time increments. We will usually omit them if it is not necessary to emphasize the time dependence of the integrands explicitly. Whenever we write $\int AdM$ this is just a short hand notation for the stochastic process $\left(\int_0^t A_s dM_s\right)_{t\geq 0}$.

Lemma 4.5 (Lemma 2.7 in [46])**.** Let $(M_t)_t$ be a martingale and consider a process $(X_t)_t \in L^2(d[M])$, i.e. its sample paths are square integrable with respect to $d[M]$. Then there exists a sequence $(X_t^{(n)})_t$ of bounded simple processes such that the pointwise limit $X_t^{(n)} \to X_t$ exists almost surely and

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^t (X_s - X_s^{(n)})^2 d[M]_s \right] = 0, \tag{4.6}$$

which expresses that the convergence also holds in $L^2(\mathcal{P})$ for all $t \geq 0$

We are now able to transfer the construction of the stochastic Itô integration to more general processes X with $X \in L^2(d[M])$.

Theorem and Definition 4.6 (compare Def. 2.9, Prop.2.10 in [46] or Thm. 3.16 in [50])**.** Let $(M_t)_t$ be a continuous square-integrable martingale and consider an adapted process [see Def. B.19] $(X_t)_t \in L^2(d[M])$. Consider a sequence of simple processes $(X_t^{(n)})_t$ approximating X

such that (4.6) holds. By taking subsequences, we can achieve that

$$\sum_n \mathbb{E}\left[\int_0^t (X_s^{(n)} - X_s)^2 d[M]_s\right]^{\frac{1}{2}} < \infty, \tag{4.7}$$

Under these assumptions, there exists a *unique square-integrable local martingale* $\left(\int_0^t X_s dM_s\right)_{t\geq 0}$, called the *Itô integral* of X with respect to M, such that for all $T \geq 0$

$$\lim_{n\to\infty} \sup_{0\leq t\leq T} \left| \int_0^t (X_s^{(n)} - X_s) dM_s \right| = 0 \tag{4.8}$$

holds almost surely and in $L^2(\mathcal{P})$.

Because the Itô integral is a (local) martingale, the expectation of the stochastic Itô integral is constant and thus identically zero. Therefore we can interpret the increment dM to have zero expectation $\mathbb{E}[dM] = 0$ in the Itô sense.

In order to define the Stratonovich integral we want to proceed analogously. However we need to ensure that the limit exists in an appropriate sense almost surely, which ultimately requires stronger assumptions on the integrand.

Theorem and Definition 4.7. Let $(M_t)_t$ be a local martingale and $(X_t)_t \in L^2(d[M])$ an adapted *semimartingale*. For a sequence $(X_t^{(n)})_t$ of simple processes approximating X in $L^2(\mathcal{P})$ satisfying the summability condition (4.7), the limit of the simple Stratonovich integtrals $\int_0^t X^{(n)} \circ dM$ exists almost surely and in $L^2(\mathcal{P})$. We then define the *Stratonovich integral* of X with respect to M to be the limiting process

$$\int_0^t X \odot dM := \lim_{n\to\infty} \int_0^t X^{(n)} \odot dM. \tag{4.9}$$

Mathematicians remained rather sceptical about Stratonovich's construction, until Itô managed to connect this approach to his own work under the premise that the integrand is a semimartingale [51]. The reason for this restriction relies on the observation, that we can express Stratonovich's integral for a simple processes $(X_t)_t$ as

$$\sum_{k=0}^{N_t} \frac{X_k + X_{k-1}}{2}(M_{t_k} - M_{t_{k-1}}) = \sum_{k=0}^{N_t} X_{k-1}(M_{t_k} - M_{t_{k-1}}) + \frac{X_k - X_{k-1}}{2}(M_{t_k} - M_{t_{k-1}}).$$

The first summand on the right hand side clearly corresponds to the Itô integral. If we consider a sequence of simple proccesses approximating a semimartingale, we see that the second summand just becomes the crossvariation of X and M when passing to the limit. If the integrand is a

semimartingale, we can therefore relate Itô and Stratonovich integration by

$$\int_0^t X \odot dM = \int_0^t X dM + \frac{1}{2}[X, M]_t. \tag{4.10}$$

In particular, this shows that the Stratonovich integral is again an adapted process. Since the expectation of the crossvariation term is not necessarily constant in time, the Stratonovich integral is generally *not* a martingale. Nevertheless it is still a *semi*martingale, since the crossvariation process has bounded first variation for all times $t \geq 0$. More details, including a complete proof of Thm. 4.7 and Eq. (4.10) can be found on p. 224f in [47], or p. 132f in [52].

Example 4.8. As an easy example, consider the case where we integrate a one-dimensional Brownian motion $(B_t)_t$ against itself. First we consider the Itô-integral. Let $T > 0$ and $B_k := B_{kt/N}$ in the following. We obtain

$$\begin{aligned}
\int_0^T B_s dB_s &= \lim_{N\to\infty} \sum_{k=1}^{N} B_{k-1}(B_k - B_k) \\
&= \frac{1}{2} \lim_{N\to\infty} \sum_{k=1}^{N} (B_k + B_{k-1})(B_k - B_{k-1}) - \frac{1}{2} \lim_{N\to\infty} \sum_{k=1}^{N} (B_k - B_{k-1})^2 \\
&= \frac{1}{2} \lim_{N\to\infty} \sum_{k=1}^{N} (B_k^2 - B_{k-1}^2) - \frac{1}{2} T = \frac{1}{2}(B_T^2 - T).
\end{aligned}$$

Because of Eq. (4.10) and the fact that $[B, B]_T = [B]_T = T$ it is easy to see that the Stratonovich integral is given by

$$\int_0^T B_s \odot dB_s = \int_0^T B_s dB_s + \frac{1}{2}[B, B]_T = \frac{1}{2} B_T^2.$$

Since the Itô integral is a square-integrable (local) martingale, it is interesting to see if its quadratic variation can be related to the integrand and the driving martingale.

Proposition 4.9 (see Prop. 2.10 in [46])**.** The quadratic variation of the stochastic Itô integral from Thm. 4.6 is given by

$$\left[\int X_s dM_s \right]_t = \int_0^t X_s^2 d[M]_s. \tag{4.11}$$

In particular, we have the following identity

$$\mathbb{E}\left[\left(\int_0^t X_s dM_s \right)^2 \right] = \int_0^t X_s^2 d[M]_s, \tag{4.12}$$

which is known as the *Itô isometry*.

The groundbreaking work of Kiyoshi Itô laid the foundations for the mathematical discipline that is known today as stochastic analysis and was highly influential on contemporary mathematicians. Notably the work of Kunita and Watanabe [53] approaches stochastic integration using the theory of Hilbert spaces. We do not want to go into this much further and only hint at the observation, that the Itô isometry (4.11) can be considered within their framework as a special case of the following identity.

Proposition 4.10 (see Prop. 2.17 in [46])**.** Let M_1, M_2 be martingales on some filtered probability space $(\Omega, \mathcal{P}, (\mathscr{F}_t)_t)$, and $X_i \in L^2(d[M_i])$ for $i = 1, 2$. Then the cross variation process of the stochatic integrals $\int X_1 dM_1, \int X_2 dM_2$ can be computed as

$$\left[\int X_1 dM_1, \int X_2 dM_2\right] = \int X_1 X_2 d[M_1, M_2]. \tag{4.13}$$

Despite having several properties which are desireable from a probabalistic point of view, the Itô integral has one major drawback. As we can already tell by the earlier Example 4.8, the fundamental theorem of calculus does no longer apply. Essentially all other inconveniences inherent to the Itô integral can be traced back to this. However, Itô noticed that only a slight alteration involving derivatives of second order is required to derive a substitute. The result is the celebrated *Itô formula*:

Theorem 4.11 (Itô formula, [47] Thm. 5.3.1)**.** Consider a local semimartingale $X = X_0 + V + M$ on $\mathbb{R}^n$ as in (4.3). Let $f \in C^{2,1}(\mathbb{R}^n \times \mathbb{R}_{\geq 0})$, i.e. f is a function tht is twice continously differentiable in the spatial component and continuously differentiable in time. Then the identity

$$\begin{aligned} f(X_t, t) = f(X_0, 0) &+ \int_0^t (\partial_s f)(X_s, s) ds + \sum_{i=1}^n \int_0^t (\partial_i f)(X_s, s) dX_s^i \\ &+ \frac{1}{2} \sum_{i,j=1}^n \int_0^t \left(\partial_i \partial_j f\right)(X_s, s) d[M^i, M^j]_s \end{aligned} \tag{4.14}$$

holds almost surely and for all times $t \geq 0$.

Heuristically it makes sense to expect that this formula involves second-order derivatives. Because the variance of Brownian motion increases linearly in time $\mathbb{E}[B_t^2] = t$, the stochastic increment can be interpreted to be of order $dB_t \approx \sqrt{dt}$. A Taylor expansion of $f(B_t)$ up to linear order in dt then involves second derivatives of f as well.

The Stratonovich integral in Example 4.8 on the other hand, resembles what we would expect from ordinary differential calculus. Indeed, with the additional the cross variation term of (4.10), we recover the ordinary chain rule.

Theorem 4.12 (Chain rule for Stratonovich)**.** Let $X = X_0 + V + M$ be an n-dimensional continuous local semimartingale as before, and $f \in C^{3,1}(\mathbb{R}^n, \mathbb{R})$. Then

$$f(X_t, t) = f(X_0, 0) + \int_0^t \partial_s f(X_s, s) ds + \sum_{i=1}^n \int_0^t (\partial_i f)(X_s, s) \odot dX_s^i. \tag{4.15}$$

holds for all $t \geq 0$ and almost surely.

Proof. In order to prove this, we actually need to take a detour via the Itô formula and the correspondence (4.10). First we apply the Itô formula and convert the appearing Itô integration to a Stratonovich integral using (4.10)

$$\begin{aligned} f(X_t, t) = \int_0^t \partial_s f(X_s, s) ds + \sum_{i=1}^n \int_0^t \partial_i f(X_s, s) \odot dX_s^i - \frac{1}{2} \sum_{i=1}^n \int_0^t d[\partial_i f(X, \cdot), M^i]_s \\ + \frac{1}{2} \sum_{i,j=1}^n \int_0^t \partial_i \partial_j f(X_s, s) d[M^i, M^j]_s. \end{aligned} \tag{4.16}$$

Combining the Itô formula with (4.13), it is easy to show that the cross variation obeys

$$[f(X, \cdot), N]_t = \int_0^t \sum_{i=1}^n (\partial_i f)(X_s, s) d[M^i, N]_s \tag{4.17}$$

for all (semi-)martingales $X = X_0 + V + M$ on $\mathbb{R}^n$, (real) martingales N and functions $f : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ which are twice continuously differentiable in the spatial components. Therefore we get

$$[\partial_i f(X, \cdot), M^i]_s = \int_0^s \sum_{j=1}^n \partial_j \partial_i f(X_r, r) d[M^i, M^j]_r. \tag{4.18}$$

If we insert (4.18) into (4.16) the last two terms cancel and we arrive at (4.15). The equation (4.18) is where we need the additional regularity $f \in C^{3,1}(\mathbb{R}^n \times \mathbb{R}_{\geq 0})$, since applying (4.17) to the left-hand side of (4.18) requires $\partial_i f$ to be twice continuously differentiable. □

When working with a Brownian motion, the assumption on the regularity of f can actually be weakened [54]. This improvement is still not sufficient to hide the fact that stochastic integration and geometry are fundamentally incompatible, and only work together if forced. In particular the chain rule we just derived for Stratonovich's integral is not an intrinsic property of this construction, but merely an imitation of the chain rule in standard analysis.

Remark 4.13. Note that the Stratonovich integral is technically not even an integral. In order to call something an integral one would expect it to satisfy estimates in terms of the zero-order information the integrand provides. It follows from (4.10) in combination with (4.17) that

$$\int_0^t f(X) \odot dM = \int_0^t f(X) dM + \frac{1}{2} f'(X_t)[X, M]_t.$$

For any estimate of the integral we can come up with, we would need to take first order derivatives of f into account as well, cf. p.225 [47].

4.2 Stochastic Differential Equations

Now that we have developed a notion of stochastic integration, we can make sense of stochastic differentiation as a inverse operation.

Definition 4.14. Let $(B_t)_t$ be a d-dimensional Brownian motion on a probability space $(\Omega, \mathcal{P}, \mathscr{F})$ and $\xi : \Omega \to \mathbb{R}^n$ a random variable. We define a filtration of $\mathscr{F}$ by $\mathscr{F}_t := \sigma(\xi; X_s \text{ for } 0 \leq s \leq t)$ which we additionally enrich with all sets of measure zero. A stochastic process $(X_t)_t$ with values in $\mathbb{R}^n$ is a *strong solution* to the Itô stochastic differential equation (SDE)

$$dX_t = a(X_t, t)dt + b(X_t, t)dB_t, \quad X_0 = \xi \tag{4.19}$$

for some measurable coefficient functions $a : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \to \mathbb{R}^n$, $b : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \to \mathrm{Mat}_{n\times d}(\mathbb{R})$, if and only if

1. X is adapted to the filtration $(\mathscr{F}_t)_t$,
2. $\mathcal{P}(X_0 = \xi) = 1$,
3. $\mathcal{P}(\int_0^t \|a(X_s, s)\| + \|b(Xs, s)\|ds < \infty) = 1$ for all $t > 0$,
4. the integral equation $X_t = X_0 + \int_0^t a(X_t, t)dt + \int_0^t b(X_t, t)dB_t$ is satisfied for all times $t \geq 0$ almost surely.

Similarly, we say that $(X_t)_t$ satisfies the Stratonovich SDE

$$dX_t = a(X_t, t) + b(X_t, t) \odot dM_t \tag{4.20}$$

if and only if $1.$-$4.$ hold, where we replace the Itô integral in $4.$ with the corresponding Stratonovich integral.

Remark 4.15. Another important tool in stochastic analysis is the concept of a *weak* solution. These are continuous processes $(X_t)_t$ such that only $3.$ and $4.$ are satisfied. The difference to a strong solution is that one leaves the freedom to choose the underlying probability space, the driving Brownian motion *and* the filtration, such that the driving Brownian motion and X are adapted.

As for classical differential equations, linear growth and Lipshitz conditions on the parameter functions guarantee existence and uniqueness of solutions to SDEs.

Theorem 4.16 (Theorem 2.9. in [46])**.** Consider a d-dimensional Brownian motion $(B_t)_t$ on some filtered probability space, and ξ a random variable with finite second moment. Let a and b be bounded measureable functions as in (4.19) which satisfy the linear growth and global Lipshitz conditions

$$\|a(t,x)\|^2 + \|b(t,x)\|^2 \leq K^2(1+\|x\|^2), \qquad \text{(linear growth condition)} \quad (4.21)$$

$$\|a(t,x) - a(t,y)\| + \|b(t,x) - b(t,y)\| \leq K\|x-y\| \qquad \text{(global Lipshitz condition)} \quad (4.22)$$

for all $t \geq 0$, $x, y \in \mathbb{R}^n$ and some constant $K > 0$. Then the Itô SDE

$$dX_t = a(X_t,t)dt + b(X_t,t)dB_t, \quad X_0 = \xi \tag{4.23}$$

admits a unique strong solution $(X_t)_t$ which is defined for all times $t \geq 0$.

Since we know how to relate Itô and Stratonovich integration, we can convert Stratonovich SDEs to Itô SDEs and vice versa if the coefficients are sufficiently nice. If the driving processes are independent one-dimensional Brownian motions, this conversion takes the following form:

Proposition 4.17. Let $B^1, \ldots, B^m$ be independent Brownian motions, $B_t := (B_t^1, \ldots, B_t^m)$ and $(X_t)_t$ a stochastic process in $\mathbb{R}^n$. Let $a : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ and $b : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \to \mathrm{Mat}_{n\times m}(\mathbb{R})$ satisfy the Lipshitz and growth conditions of Thm. 4.16. We assume that the matrix coefficient $b = (b_1, \ldots, b_n) = (b_{ij})_{i\leq n, j\leq m}$ is continuously differentiable in the spatial components. Then the process $(X_t)_t$ satisfies the Itô SDE

$$dX_t = a(X_t,t)dt + b(X_t,t)dB_t \tag{4.24}$$

if and only if it satisfies the Stratonovich SDE

$$dX_t = \left(a(X_t,t) - \frac{1}{2}c(X_t,t)\right)dt + b(X_t,t) \odot dB_t, \tag{4.25}$$

where the drift correction is given by

$$c_j(x,t) := \sum_{i=1}^{n}\sum_{k=1}^{m}\left(\frac{\partial}{\partial x_i}b_{jk}(x,t)\right)b_{ik}(x,t). \tag{4.26}$$

Proof. The starting point to prove this claim is the identity (4.10). Let X be the solution to the Itô SDE (4.24), i.e.

$$X_t = X_0 + \int_0^t a(X_s,s)ds + \int_0^t b(X_s,s)dB_s \quad \text{for all } t \geq 0.$$

With the understanding that stochastic integration is performed in each coordinate seperately and that $b(X_s,s)dB_s = \sum_i b_i(X_s,s)dB_s^i$, we can convert the second integral into a Stratonovich

integral using Eq. (4.10) and get

$$X_t = X_0 + \int_0^t a(X_s, s)ds + \int_0^t b(X_s, s) \odot dB_s - \frac{1}{2}c(X_t, t). \tag{4.27}$$

for all $t \geq 0$ where $c = (c_1, \ldots, c_n)$ with $c_j(X_s, s) := \sum_k [b_{jk}(X_\cdot, \cdot), B^k]_s$. With Itô's formula (4.11), the identity (4.13), and the fact that the crossvariation with terms of bounded variation vanishes, we can compute the correction term explicitly as

$$\begin{aligned} c_j(X_t, t) &= \sum_k \left[b_{jk}(X, \cdot), B^k\right]_t = \sum_k \left[\int \sum_i \partial_i b_{jk}(X_s, s) dX_s^i, B^k\right]_t \\ &= \sum_{k,i} \int_0^t \partial_i b_{jk}(X_s, s) d[X^i, B^k]_s = \sum_{k,l,i} \int_0^t \partial_i b_{jk}(X_s, s) d\left[\int b_{il}(X_r, r) dB_r^l, B^k\right]_s \\ &= \sum_{i,l,k} \int_0^t \left(\partial_i b_{jk}(X_s, s)\right) b_{il} d[B^l, B^k]_s = \sum_{i,k} \int_0^t \left(\partial_i b_{jk}(X_s, s)\right) b_{ik} ds \end{aligned}$$

which is exactly as claimed in Eq. (4.26). Starting with a Stratonovich SDE, we can work the argumentation backwards analogously, which concludes the proof. □

4.3 Itô versus Stratonovich

In the following, we compare how the two notions of stochastic integration behave under different aspects. As far as stochastic properties are concerened, we already saw that the Itô integral is a (local) martingale and therefore very much suited for the framework of stochastic analysis.

Underlying Geometry

Another question of interest is concerned with the geometry of the underlying space: Assuming that the stochastic increment is locally tangent to a submanifold of the euclidean space, does the process stay on this submanifold? This is fundamentally relevant for our purposes, as we ultimately want to model SDEs on Lie groups. For Stratonovich's integral the answer is generally affirmative, since the chain rule is satisfied.

Proposition 4.18. Consider an embedded m-dimensional submanifold $\mathbb{M} \subset \mathbb{R}^d$ (without boundary). Let $V, W_1, \ldots, W_k$ be vector fields on $\mathbb{R}^d$ which satisfy the growth and Lipshitz conditions, and restrict to vector fields on the submanifold $\mathbb{M}$. For real-valued Brownian motions $B^1, \ldots, B^k$, the process defined as the solution of the Stratonovich SDE

$$dX_t = V(X_t)dt + \sum_i W_i(X_t) \odot dB_t^i. \tag{4.28}$$

stays on $\mathbb{M}$ for all times if initially $X_0 \in \mathbb{M}$.

Proof. Note that it suffices to check the statement locally. Since $\mathbb{M}$ is an embedded submanifold we can choose a local chart φ on an open neighbourhood $U \subset \mathbb{M}$ around $p := X_0 \in \mathbb{M}$. Because the vector fields are tangent to $\mathbb{M}$, we can consider the localised SDE

$$d\widetilde{X}_t = V^\varphi(\widetilde{X})dt + \sum_i W_i^\varphi(\widetilde{X}_t) \odot dB_t^i, \quad \widetilde{X}_0 = \varphi(p) \tag{4.29}$$

where $V^\varphi(x) := D\varphi V(\varphi^{-1}(x))$ denotes the vector field corresponding to V in the coordinate representation, similarly for the W_i. The vector fields $V^\varphi, W_1^\varphi, \dots, W_k^\varphi$ satisfy Lipshitz and growth conditions at least on some small precompact neighbourhood around $\varphi(p)$. Therefore there exists a stopping time T with $T > 0$ almost surely such that the solution $(\widetilde{X}_t)_{t \le T}$ to (4.29) with $\widetilde{X}_0 = \varphi(p)$ is defined and stays inside $\varphi(U)$. If we map the solution $(\widetilde{X}_t)_{t \le T}$ back to $\mathbb{M}$ and apply the chain rule once more, we see that $X_t := \varphi^{-1}(\widetilde{X}_t) \in \mathbb{M}$ solves the SDE (4.28) for $t \le T$. Because our assumptions gurantee the uniqueness of the solution, this concludes the proof. □

In general this procedure will fail when working with Itô SDEs. This does not come as a surprise, considering the behaviour of the Itô increment under non-linear coordinate transformations. The following simple example supports this intuition:

Example 4.19. Consider the Itô SDE

$$(dX_t, dY_t) = (-Y_t, X_t)dB_t, \tag{4.30}$$

where B_t is a one-dimensional Brownian motion. The stochastic increment $(-Y_t, X_t) \in T_{(X_t,Y_t)}\mathbb{S}^1$ is tangent to the unit sphere. Let $f(x, y) = x^2 + y^2$ and assume that initially $(X_0, Y_0) \in \mathbb{S}^1$. If the process stays on $\mathbb{S}^1$ we would nececcarily have $df(X_t, Y_t) = 0$. Because the mixed derivatives $\partial_x \partial_y f = 0$ vanish, using both Itô's formula and isometry, we compute

$$\begin{aligned} df(X_t, Y_t) &= \underbrace{2X_t dX_t + 2Y_t dY_t}_{=0} + \frac{1}{2}\left(\partial_x^2 f(X_t, Y_t) d[X]_t + \partial_y^2 f(X_t, Y_t) d[Y]_t\right) \\ &= d[X]_t + d[Y]_t = d\left[\int -Y_s dB_s\right]_t + d\left[\int X_s dB_s\right]_t \\ &= d\int_0^t Y_s^2 + X_s^2 ds = (X_t^2 + Y_t^2)dt. \end{aligned}$$

In particular, $f(X_t, Y_t)$ is monotonically increasing and $df(X_t, Y_t)$ is non-zero for all times if initially $(X_0, Y_0) \in \mathbb{S}^1$. Therefore the process will not stay on $\mathbb{S}^1$ for any $t > 0$.

This example is nicely demonstrated in Fig. 4.1 by a simulation using the Euler-Maruyama scheme, which we will briefly discuss at the end of this chapter in Proposition 4.24.

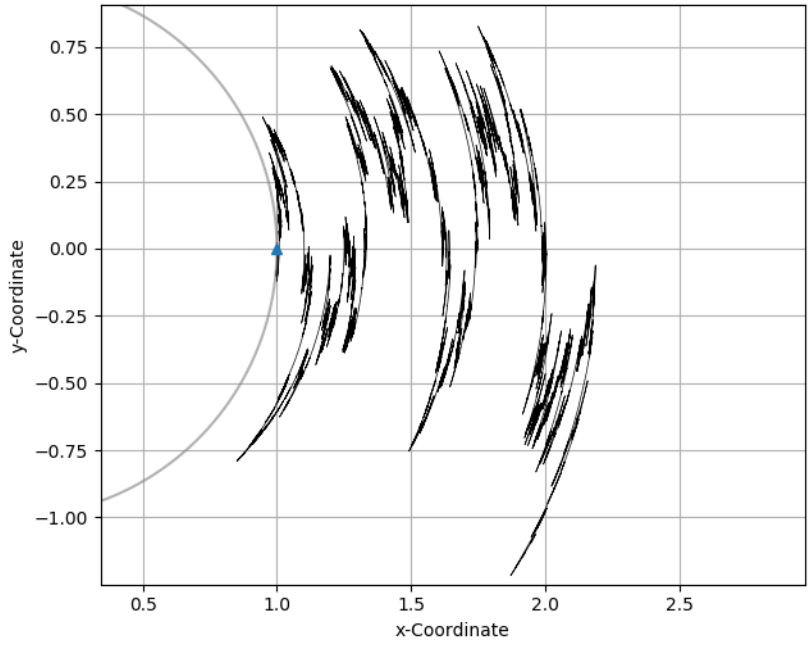

Fig. 4.1: Sample path (black) of an Itô-solution to $d(X_t, Y_t) = 0.4(Y_t, -X_t)dB_t$ starting at $(1, 0) \in \mathbb{S}^1$ (grey) with 10000 iterations and time-step $\Delta t = 0.001$ using the Euler-Maruyama scheme, Prop. 4.24. The choice of the factor 0.4 is purely for aesthetical purposes. We clearly see that the plotted sample path is contantly moving away from the unit circle even though the increment is tangent to it.

Approximations

Another major concern is the validity of different kinds of approximations. Assume that we have a Brownian motion B and that we want to physically reproduce an evolution that is simulated by the formal Langevin equation

$$\frac{d}{dt}X_t = a(X_t, t) + b(X_t, t)\xi_t \tag{4.31}$$

where ξ is distributed according to the white noise measure. A priori it is not clear if we should understand the formal increment $\xi_t = dB_t/dt$ in the Itô or Stratonovich sense. Since we cannot physically simulate distributional pulses, we need to rely on approximations to the increment dB_t/dt in one way or another. In order to produce meaningful results, the solution to the idealised equation (4.31) needs to resemble the solution of the approximated equation. Here again, Itô and Stratonovich increments behave rather differently.

Example 4.20. Consider a smooth function $f \in C^\infty(\mathbb{R}_{\geq 0} \times \mathbb{R})$ for which the second spatial derivative $\partial_x^2 f$ is not identically zero and consider the Itô increment

$$dX_t = (\partial_t f)(t, B_t)dt + (\partial_x f)(t, B_t)dB_t, \quad X_0 = f(0,0) \tag{4.32}$$

for a Brownian motion B. In order to solve this, we could naively approximate the driving Brownian motion locally uniformly via (piecewise) smooth processes $(B_t^n)_t$ satisfying $B_0^n = 0$, compute the solutions X^n to the equations

$$dX_t^n = \partial_t f(t, B_t^n) + \partial_x f(t, B_t^n)dB_t^n, \quad X_0^n = f(0,0), \tag{4.33}$$

and obtain the solution to (4.32) via the limit $X_t := \lim_{n\to\infty} X_t^n$. Because the B_t^n are differentiable, we can rewrite

$$dX_t^n = \partial_t f(t, B_t^n) + \partial_x f(t, B_t^n)\dot{B}_t^n dt. \tag{4.34}$$

Using the chain rule of ordinary calculus, we see that (4.34) is solved by $X_t^n = f(t, B_t^n)$ which converges to $f(t, B_t)$ as $n \to \infty$. But if we compute $df(t, B_t)$ using the Itô formula, we additionally get the Itô correction term $\frac{1}{2}\partial_x^2 f(t, B_t)dt$! That is, the Itô integral is in general unstable under approximations of the noise process, and the described procedure will in general fail to solve the SDE considered! The problem with this procedure is that the quadratic variation is not stable under approximations. For piecewise smooth processes approximating the Brownian motion, the quadratic variation will be identically zero, whereas it grows linearly in time for a Brownian motion.

This of course has consequences for how we need to model physical systems. In any experimental setup, the driving white noise is only approximated by piecewise constant random pulses. Interpreting (4.31) as an Itô SDE would therefore produce results not in accordance to what is observed in the system. Luckily, the Stratonovich integral behaves much better in this regard, as the following two theorems due to Stroock and Varadhan show.

Theorem 4.21 (Theorem 4.1 in [55])**.** Let $b : \mathbb{R}_{\geq 0} \times \mathbb{R}^m \to \mathrm{Mat}_{m\times d}(\mathbb{R})$ a smooth map and $a : \mathbb{R}_{\geq 0} \times \mathbb{R}^m \to \mathbb{R}^m$ a time-dependent vector field. Consider a standard Brownian motion $(B_t)_t$ on $\mathbb{R}^d$, which we approximate by a sequence of piecewise affine linear functions

$$\xi^{(n)}(t) := B(t_-^{(n)}) + \frac{t - t_-^{(n)}}{n}\left(B(t_+^{(n)}) - B(t_-^{(n)})\right)$$

with $t_\pm^{(n)} \in \frac{1}{n}\mathbb{N}$ such that $t_-^{(n)} \leq t < t_+^{(n)}$. Let $x_0 \in \mathbb{R}^m$ and consider the family of solutions $(X_t^{(n)})_t$ to the regular random integral equations

$$X_t^{(n)} = x_0 + \int_0^t a(X_s^{(n)}, s)ds + \int_0^t b(X_s^{(n)}, s)\xi_s^{(n)}ds \tag{4.35}$$

and the Stratonovich solution $(X_t)_t$ to

$$X_t = x_0 + \int_0^t a(X_s, s)ds + b(X_s, s) \odot dB_s. \tag{4.36}$$

Then $P^{(n)} = Law(X^{(n)})$ converges weakly to $P = Law(X)$ as n goes to ∞.

The expression $b(x,t) \odot dB_t$ is just a compact way of writing $\sum_i b_i(x,t) \odot dB_t^i$ for some time-dependent vector fields $b_1(x,t), \ldots, b_d(x,t)$ and independent Brownian motions $B^1, \ldots, B^d$. In the language of control systems, the theorem implies that the law $\mu_t = Law(X_t)$ of the solution is supported in the set attainable by the control system $(a|b_1, \ldots, b_d)$ at time t from x_0, i.e.

$$\mathrm{supp}(\mu_t) \subset \mathcal{A}_t(x_0).$$

Another major accomplishment of Stroock and Varadhan was to show that the converse inclusion also holds, which is implied by the next theorem.

Theorem 4.22 (Theorem 5.1 in [55], related to Theorem 9 in [56])**.** Consider a, b, B as in the previous Theorem 4.21 and a process $(\eta_t)_t$ adapted to $(B_t)_t$ with twice continuously differentiable sample paths. Let X be a solution to the Stratanovich SDE (4.36) and X^η a solution to the random integral equation

$$X_t = X_0 + \int_0^t a(X_s, s)ds + \int_0^t b(X_s, s)d\eta_s.$$

Then the probability that the solutions X and X^η differ at most by $\varepsilon > 0$ is controlled by the difference $\eta - B$ in the sense that the conditional probability satisfies

$$\lim_{\delta \to 0_+} \mathcal{P}\left(\sup_{0 \le t \le T} ||X_t - X_t^\eta|| < \varepsilon \middle| \sup_{0 \le t \le T} ||\eta_t - B_t|| < \delta \right) = 1. \tag{4.37}$$

Corollary 4.23 (part of Theorem 8.3.5 in [47])**.** Let $V, W_1, \ldots, W_m$ be vector fields on $\mathbb{R}^d$ and $B^1, \ldots, B^m$ independent real-valued Brownian motions. Consider the solution to the Stratonovich SDE

$$dX_t = V(X_t) + \sum_j W_j(X_t) \odot dB_t^j, X_0 = x_0.$$

Then the support of $\mu_t = Law(X_t)$ is equal to the set which is attainable by the control system $(V|W_1, \ldots, W_m)$ from x_0 at time t.

Thanks to this this corollary, Thm. 4.22 is usually referred to as the *Support Theorem*.

It is important to note that in both cases the somewhat restricted form of approximation to the noise is required. This is linked to the fact that linear maps are generally non-commutative if we consider more than one dimension, which can cause severe problems in such approximations. An

example of an ODE where the solutions of the approximation do not converge to the true solution can be found for instance on p. 38 of [56].

When numerically solving stochastic differential equations, one can come up with another conceptually different type of approximation. In the spirit of the Euler method for ordinary differential equations, we want to solve a SDE using Taylor expansion and extrapolate the solution over small time intervals by

$$X_{t+\Delta t} := X_t + a(X_t, t)\Delta t + b(X_t, t)(B_{t+\Delta t} - B_t). \tag{4.38}$$

This procedure is known as the *Euler-Maruyama scheme*. By default this procedure favours Itô's integral over Stratonovich's. This is because we extrapolate using the currently known value, which corresponds to the lower boundary of the interval we consider. Formally, this is captured by the following theorem.

Proposition 4.24 (see Theorem 10.2.2 of [57])**.** Consider an Itô stochastic differential equation

$$dX_t = a(X_t, t)dt + b(X_t, t)dB_t,\ X_0 = x_0.$$

where the coefficients satisfy the Lipshitz and growth condition of Thm. 4.16 and assume additionally that

1. $\mathbb{E}[\|X_0\|^2] < \infty$,
2. $\|a(t,x) - a(t,y)\| + \|b(t,x) - b(t,y)\| \leq C_1\|x - y\|$,
3. $\|a(t,x)\| + \|b(t,x)\| \leq C_2\|x - y\|$,
4. $\|a(s,x) - a(t,x)\| + \|b(s,x) - b(t,x)\| \leq C_3(1 + \|x\|)|s - t|^{1/2}$,

hold for suitable constants $C_1, C_2, C_3 > 0$. For $\delta > 0$, we consider the discrete approximation X^δ constructed iteratively by

$$X^\delta_{t_{k+1}} := X^\delta_{t_k} + a(X^\delta_{t_k}, t_k)\delta + b(X^\delta_{t_k}, t_k)B_\delta$$

with $t_k = k\delta$ and linearly interpolating between the anchor points. Then for every $T > 0$, there exists a constant $C > 0$ such that the difference between the true solution X and the approximation X^δ at T can be bounded in expectation by

$$\mathbb{E}[\|X_T - X^\delta_T\|] \leq C\delta^{1/2}.$$

It is evident that such a procedure must fail for the Stratonovich integration. In order to obtain a Stratonovich solution we would need to replace $b(X_k, t_k)$ in the extrapolation (4.38) by

$b\left(X_{k+\delta/2}, \frac{(t_{k+1}+t_k)}{2}\right)$. But of course at time t_k, we cannot determine the distribution of X^δ for $(t_k + t_{k+1})/2$ yet.

Conclusion

We conclude this capter by summarsing how Itô and Stratonovich integrals behave differently under several aspects.

Itô Calculus	**Stratonovich Calculus**
+ Stochastic increment has expectation zero	- Expectation of stochastic increment
+ Itô Isometry (Prop 4.9)	is nonzero in general
- Unstable under approximation of	+ Stable under approximations of
the noise term (Ex. 4.20)	the noise (Thm. 4.21)
- Ordinary chain rule does not apply (Thm. 4.11)	+ Chain rule holds as usual, (Prop 4.12)
- Does not respect underlying geometry (Ex. 4.19)	+ Preserves geometry of underlying
+ Euler-Mayurama scheme for solving	submanifolds (Prop 4.18)
SDEs (Prop 4.24)	

The Itô calculus is designed to behave well for probabalistic computations and mathematical models, while Stratonovich's notion of stochastic integration features properties which are more desireable from a physical perspective. In particular it is the adequate notion to interpret formal Langevin-type equations

$$\frac{d}{dt}X_t = a(X_t, t) + b(X_t, t)\xi_t \tag{4.39}$$

which are driven by white noise ξ, the distributional derivative of Brownian motion. Most importantly, the two developed notions of stochastic integration complement each other nicely, and as we may convert them back and forth for sufficiently nice integrands, we can exploit the respective advantages of both and avoid the involved difficulties.

5. LÉVY PROCESSES ON LIE GROUPS

In this chapter we discuss properties of stochastic processes on Lie groups that naturally arise by geometric integration of Lévy processes in (matrix) Lie algebras. We use the results of the previous discussion about stochastic integration in order to give a proper infinitesimal descriptions of such processes. As a preperation for Part III, we are primarily concerned with their Markov properties and the generators of their semigroups. From now on, G is always a connected Lie group with identity element 1_G and $\mathfrak{g} = T_{1_G}G$ its Lie algebra, unless indicated differently. We write $X = x^l$ for the left invariant vector field corresponding to $x \in \mathfrak{g}$. In order to avoid the treatment of stochastic calculus on manifolds, we usually assume G to be a matrix Lie group, i.e. given by some faithful unitary representation such that $G \subset \mathrm{GL}(d)$, and consequently $\mathfrak{g} \subset \mathfrak{gl}_d$. The left-invariant vector field X corresponding to $x \in \mathfrak{g}$ is simply given by $X(U) = Ux \in \mathrm{End}(\mathbb{C}^d)$ in this case.
As usual, $(\Omega, \mathcal{P}, (\mathscr{F}_t)_{t\geq 0})$ denotes the filtered probability space on which we model all stochastic processes. In particular, we consider all occuring processes to be adapted to the filtration $(\mathscr{F}_t)_t$ (Def. B.19). All densities and the occuring L^p-spaces are to be understood with respect to the Haar measure. As general references of this chapter, we refer to [58, 59, 20, 60].

5.1 Stochastically Controlled Systems

The obvious procedure to generate random distributions on a (matrix) Lie group G using an invariant control system $(x|y_1, \ldots, y_m)$ is to apply random pulses to the controlled parts. Specifically, we assume that the random control functions are simple processes of the form

$$c(t) = \sum_{i\geq 0} a_i \chi_{[t_i, t_{i+1})}(t),$$

with $t_i = i\Delta T$ and independent random coefficients $a_i \sim \mathcal{N}(0, \Delta T^{-1})$ for some time step $\Delta T > 0$. As usual, χ denotes the characteristing function of the respective time intervals. We choose such processes c_j for each control y_j independently and steer the system via the random differential equation

$$\frac{d}{dt}U_t = U_t\big(x + \sum_j c_j(t) y_j\big), \quad U_0 = 1_G. \tag{5.1}$$

The control parameters c_j are chosen such that they are the derivatives of piecewise linear approximations to independent Brownian motions $B^1, \dots, B^m$ in the spirit of Prop. 4.21. This proposition implies that the process $(U_t)_t$ in (5.1) corresponds in the limit of $\Delta T \to 0$ to the solution of the *Stratonovich* stochastic differential equation

$$dU_t = U_t x dt + \sum_j U_t y_j \odot dB_t^j, \quad U_0 = 1_G. \tag{5.2}$$

The SDE (5.2) indeed describes an evolution on a Lie group since, the Stratonovich integration respects the underlying geometry. We also see that (5.2) defines a continuous process, since both the driving noise as well as the coefficient vector fields are continuous. For a probabalistic treatment the Itô form of an SDE is preferrable over the Stratonovich version, following the discussion of Chapter 4. If we consider $G \subset \mathrm{SU}(d)$ in its standard representation, we can use the Itô-Stratonovich conversion (Prop. 4.17, to derive the corresponding Itô SDE

$$dU_t = U_t \left(x + \frac{1}{2} \sum_j y_j^2 \right) dt + \sum_j U_t y_j dB_t^j, \quad U_0 = 1_G, \tag{5.3}$$

where we understand the muliplication y_j^2 in terms of endomorphisms on $\mathbb{C}^d$. Both the linear growth and the Lipshitz condition of Thm. 4.16 are satisfied, which ensures uniqueness and existence of the solution to (5.3) (and thereby also for (5.2)) for all times $t \geq 0$. Because of the underlying group structure, solving the Stratonovich SDE (5.2) amounts to a geometric integration of the process $(Z_t)_t$ on $\mathfrak{g}$ defined by $Z_t = xt + \sum_j y_j B^j$, since we can rewrite (5.2) as

$$dU_t = U_t \odot dZ_t. \tag{5.4}$$

Since the Hamiltonian is constant over small time intervals, the solution $(U_t^{(n)})_t$ to the random differential equation (5.1) with timestep $\Delta t = \frac{1}{n}$ can be explicitly stated as

$$U_t^{(n)} = \left(\prod_{j=1}^{N_t} \exp\left(Z_{j/n} - Z_{(j-1)/n} \right) \right) \exp(Z_t - Z_{N_t/n}) \tag{5.5}$$

with $N_t = \max\{j \in \mathbb{N} : j/n \leq t\}$. Here we consider the terms in the product to be multiplied from the left to the right. Because $U_t^{(n)}$ converges to U_t in distribution, we can express the distribution of U_t as the weak limit

$$U_t \overset{\mathcal{D}}{=} \lim_{n \to \infty} \prod_{k=1}^{N_t} \exp\left(Z_{k/n} - Z_{(k-1)/n} \right). \tag{5.6}$$

Recall that a Lévy process in euclidean space is a process with independent and stationary increments. In particular this is the case for the process Z above, which is essentially a Brownian

motion with an additional drift in the Lie algebra $\mathfrak{g}$. It is straightforward to see that the stationarity and independence of the increments of Z carry over to the *right* increments of $(U_t)_t$, in accordance to the order of multiplication in (5.5). Therefore it seems adequate to develop a corresponding notion of Lévy processes on Lie groups.

5.2 Lévy Processes on Lie Groups

Before we actually start, recall that we define the convolution of two measures μ, ν on G by

$$\mu * \nu(A) = \int \chi_A(gh) d\mu(g) d\nu(h).$$

Similarly we define the convolution of measurable functions f_1, f_2 with respect to the Haar measure via

$$f_1 * f_2(h) = \int f_1(hg^{-1}) f_2(g) d\mu_G.$$

Definition 5.1. Let G be a Lie group. A stochastic process $(g_t)_{t\geq 0}$ on G is called a *left Lévy process* if

1. $(g_t)_t$ has càdlàg sample paths (see Def. B.20),
2. the right increments $g_0^{-1}g_1, \ldots, g_{n-1}^{-1}g_n$ are independent for all $n \in \mathbb{N}$ and $t_1 < \cdots < t_n$,
3. the right increments are stationary, i.e. $g_s^{-1}g_t \sim g_0^{-1}g_{t-s}$ have the same law for $t \geq s$.

Similarly, we can define right-Lévy processes using the left-increments. At first it seems somewhat counterintuitive to call the process above *left* Lévy process if the *right* increments are stationary and vice versa, the reason for this is becomes clearer when looking at their Markov properties. In the following, Lévy processes should *always* to be understood as *left* Lévy processes unless indicated otherwise. The competing notations $(g_t)_t$ and $(U_t)_t$, which we are employing for Lévy processes, are supposed to indicate whether results are of a general nature or refer to the setting we defined in Eq. (5.2) respectively.

For a Lévy process $(g_t)_t$, we will denote its law at time t by μ_t and write $\mu_t^0 = Law(g_0^{-1}g_t)$ for the law of the process starting at the identity. We begin with two important observations.

1. Since the paths of a Lévy process are càdlàg, they are right continuous and therefore we have $\lim_{t\to 0} \mu_t^0(U) = 1$ for every open neighbourhood U around 1_G.
2. The independence and stationarity of the increments manifest themselves in the observation that $\mu_{t+s} = \mu_t * \mu_s^0 = \mu_s * \mu_t^0$ holds for all $s, t \geq 0$.

With the initial condition $g_0 = 1_G$ we therefore have $\mu_{t+s} = \mu_t * \mu_s$ for every $s, t \geq 0$ and $\lim_{t\to 0} \mu_t = \delta_{1_G}$ weakly which gives the family $\{\mu_t\}_{t\geq 0}$ the structure of a *continuous convolution semigroup*.

5.3 Markov Properties and Generators

As in the euclidean case, Lévy processes on Lie groups are a special instance of time-homogenous Markov processes. This is important, because Markov processes are 'memoryless', and in order to get a uniform distribution after always starting at the same point, we need to 'forget' this information. The Markov property of a Lévy process $(g_t)_t$ follows directly from the fact that the increment $g_s^{-1}g_t$ for $t \geq s$ is independent of $\mathscr{F}_s$, and the time homogenity corresponds to the increments being stationary. Therefore the transition operators $P_{t,s}$, which are defined on measureable functions by

$$P_{t,s}f(h) := \mathbb{E}[f(g_t)|g_s](h) = \mathbb{E}[f(g_s g_s^{-1} g_t)|g_s = h] = \mathbb{E}[f(hg_0^{-1}g_{t-s})],$$

depend only on the difference $t - s$. In particular we have $P_{0,0} = \mathbb{1}$ and the family $\{P_t\}_t$ with $P_t := P_{t,0}$ inherits the structure of an operator semigroup, according to the Chapman-Kolmogorov equations (cf. Def. B.21). With the earlier observation that $\mu_{t+s} = \mu_t * \mu_s^0$ for all $s, t \geq 0$ we see that the transition kernels $\{\pi_t\}_t$ of the process simply act on probability measures $\nu \in \mathcal{M}^1(G)$ by convolutions, $\nu^{\pi_t} = \nu * \mu_t^0$. It is a standard result in the theory of Markov processes, that $(g_t)_t$ is probabalistically completely determined by each of the structures $\{P_t\}_t$, $\{\pi_t\}_t$ up to an initial condition $g_0 \sim \mu_0$, cf. Chapter 1 in [20]. In the remainder of this section we will be mostly concerned with discussing the transition operator semigroup $\{P_t\}_t$ acting on the Banach space $C_0(G)$ of continuous functions vanishing at ∞. We continue by collecting some more properties of $\{P_t\}_t$.

Proposition 5.2. The operator semigroup $\{P_t\}_t$ of a left-Lévy process acting on $C_0(G)$ is a *probability semigroup*. That is, additionally to the semigroup properties, it satisfies the following conditions for all times $t \geq 0$ and $f \in C_0(G)$ (compare Def. 3.4 in [61]):

1. $P_t f \geq 0$ if f is nonnegative *(positivity)*.
2. $\lim_{t\to\infty} \|P_t f - f\|_\infty = 0$ *strong continuity*.
3. $P_t 1 = 1$ for all $t \geq 0$, if G is compact. *(mass-preserving)*
   Otherwise, there exists a sequence f_n such that $\sup \|f_n\|_\infty < \infty$ and $P_t f_n \to 1$ pointwise as $n \to \infty$.
4. $\|P_t f\|_\infty \leq \|f\|_\infty$ *(contracitvity)*.

Besides, we also have that P_t commutes with left-multiplication maps $l_h : G \to G$

5. $P_t l_h^* f = l_h^* P_t f$ for all $f \in C_0(G)$ (*left-invariance)*,

which is the particular reason we call the process a *left* Lévy process.

Proof. Because $P_t f(g) = \int_G f(hg) d\mu_t(h)$, Claims $4, 1$ and the compact case of 3 follow immediately. For the non-compact case in 3, we choose a sequence of functions $f_n \in C_0(G)$ with values

in $[0,1]$ such that $f_n \to 1$ pointwise and $f_n^{-1}(\{1\}) \subset f_m^{-1}(\{1\})$ for $n \leq m$. Left-invariance is esily verified by

$$P_t l_h^* f(g) = \mathbb{E}[f(hgg_0^{-1}g_t)] = P_t f(hg) = l_h^* P_t f(g).$$

We are only left to show strong continuity in ∞-norm. Fix $\varepsilon > 0$ and assume that G is compact. Because f is continuous, we find a neighbourhood U_ε around 1_G such that $|f(g) - f(gh)| < \varepsilon$ for all $h \in U_\varepsilon$ and $g \in G$. Because the convolution semigroup $\{\mu_t^0\}_t$ is continuous, there exists a time $t > 0$ such that $\mu_s(G \setminus U_\varepsilon) < \varepsilon$ for all $s \leq t$. This allows to derive the estimate

$$\begin{aligned} \|P_s f - f\|_\infty &\leq \sup_g \int_{U_\varepsilon} |f(gh) - f(g)| d\mu_s(h) + \sup_g \int_{G \setminus U_\varepsilon} |f(gh) - f(g)| d\mu_s(h) \\ &\leq (1 + \|f\|_\infty)\varepsilon. \end{aligned} \tag{5.7}$$

If G is not compact, we can still find a compact $K \subset G$ such that $|f| \leq \varepsilon/2$ outside K. Consider another compact set K' containing K such that there is an intermediate open set V with $K \subset V \subset K'$ and replace G by K' in the argument above. The intermediate open set ensures that we can shrink U_ε to achieve $g \cdot U_\varepsilon \cap K = \emptyset$ for all $g \in G \setminus K'$ if necessary. Then the estimate (5.7) holds analogously. □

Combining the semigroup property with the strong continuity, we can completely recover the semigroup only by knowing the operators for a sequence of times converging to zero. Thus the semigroup is completely determined by its ininitesimal behaviour at $t = 0$. The best way to capture this properly is to consider its *infinitesimal generator*, which is the operator $\mathcal{L}$ defined by

$$\mathcal{L}f := \lim_{t \to 0} \frac{1}{t}(P_t f - f) \tag{5.8}$$

for all $f \in C_0(G)$ where the limit exists. The limit (5.8) in fact exists on a dense subset of $C_0(G)$ as the next proposition shows.

Proposition 5.3 (cf. Thm. 13.35 in [62], and p. 304f in [63])**.** For a semigroup of the probability operators $\{P_t\}_t$ acting on $C_0(G)$, the infinitesimal generator $\mathcal{L}$ defined in Eq. (5.8) has a dense domain of definition $\mathrm{Dom}(\mathcal{L}) \subset C_0(G)$. Additionally, the semigroup preserves the domain in the sense that $P_t f \in \mathrm{Dom}(\mathcal{L})$ for all $f \in \mathrm{Dom}(\mathcal{L})$ and $t \geq 0$.

Using the definition of the infinitesimal generator, we can easily derive a differential equation describing the operator semigroup $\{P_t\}_t$ in terms of the generator $\mathcal{L}$ by

$$\partial_t P_t f = \lim_{s \to 0} s^{-1} \left(P_{t+s} - P_t\right) f = P_t \lim_{s \to 0} s^{-1}(P_s f - f) = P_t \mathcal{L} f. \tag{5.9}$$

This is a special case of the more general *Kolmogorov forward equation*. The respective *backwards*-equation is $\partial_t P_t f = \mathcal{L} P_t f$ (also known as the *Heat equation*). In order to appreciate this choice of naming properly one needs to look at the corrsponding equations in the time-*inhomogenous*

case, where $\mathcal{L}$ is time-dependent and locally characterises the changes of P_t, see for instance in [61]. In view of (5.9), one is tempted to express the semigroup $\{P_t\}_t$ via the operator exponentials $P_t = \exp(t\mathcal{L})$. The problem here is that the operator $\mathcal{L}$ is unbounded and in general *not* self-adjoint, thus we cannot make sense of this exponential. This can be circumvented by considering the defining approximation to $\mathcal{L}$ by the bounded operators $\mathcal{L}_\varepsilon := \varepsilon^{-1}(P_\varepsilon f - f)$. With this trick we can recover the semigroup from the generator via the assignment $P_t f := \lim_{\varepsilon\to 0} \exp(t\mathcal{L}_\varepsilon)f$ for all $t \geq 0$ and $f \in \mathrm{Dom}(\mathcal{L})$. A formal proof of this is given in Thm. 13.35 of [62], thereby showing that the probability semigroup and its generator determine each other uniquely.

Of course such generators can be defined for more general Markov processes whenever the transition operators are strongly continuous. For the particular case of Lévy processes they are very well understood. In the euclidean setting, this is accomplished by the *Lévy-Khinchin formula*, see for instance Chapter 2 in [47]. The result of Hunt [64] generalises this to the setting of Lie groups.

Theorem 5.4 (Hunts formula, Thm. 5.1 in [64], Thm .1.1 in [59])**.** Let $(g_t)_t$ be a Lévy process on a Lie group G and $x_1, \ldots, x_n$ a basis for its Lie algebra $\mathfrak{g}$. As ususal let $X_1, \ldots X_n$ denote the respective left-invariant vector fields. Consider a collection of compactly supported functions $y_1, \ldots, y_n$ defining a local chart around 1_G which satisfy $x_i(y_j) = \delta_{ij}$. Then the generator $\mathcal{L}$ of $(g_t)_t$ acting on smooth functions $f \in C^\infty(G)$ can expressed as

$$\begin{aligned}\mathcal{L}f(g) =& \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}X_iX_jf(g) + \sum_{i=1}^{n} b_iX_if(g) \\ &+ \int_G \left(f(gh) - f(g) - \sum_{i=1}^{n} y_i(h)X_if(g) \right) d\Pi(h), \end{aligned} \tag{5.10}$$

where $A = (a_{ij})_{i,j}$ is a positive semidefinite matrix, $b = (b_1, \ldots, b_n)$ are constants and Π a measure on G satisfying $\Pi(\{1_G\}) = 0, \int \sum_i y_i^2 d\Pi < \infty$, and $\Pi(G \setminus U) < \infty$ for every open neighbourhood U around 1_G. The measure Π is called the *Lévy measure* of the process. Conversely, every generator of the form (5.10) for such a triple (A, b, Π) defines a Lévy process on G.

Remark 5.5. The Lévy measure Π is the intensity of a pure jump process resembling the occuring discontinuities of the Lévy process. In particular, it is identically zero if the process is continuous, which significantly simplifies the expression (5.10). Since this is the case we are eventually interested in, we assume all Lévy processes to be *continuous* from now on. Even though we only state the formula (5.10) for $f \in C_0^\infty(G)$, which is strictly smaller than the domain of $\mathcal{L}$, this determines $\mathcal{L}$ completely. Here we rely on the fact (Theorem 13.35.8, Chapter II in [65]) that C_0^∞ is a *core* for $\mathcal{L}$, i.e. the domain of $\mathcal{L}$ is the closure of $C_0^\infty(G)$ with respect to the graph norm $\|f\|_{\mathcal{L}} = \|f\| + \|\mathcal{L}f\|$.

If we go back to the process (5.2) from the beginning, we expect that the infinitesimal generator

$\mathcal{L}$ of the generated Lévy process can be expressed in terms of the control and drift vector fields of the underlying control system. Consider the solution $(U_t)_t$ to the Stratonovich SDE

$$dU_t = X(U_t)dt + \sum_j Y_j(U_t) \odot dB_t^j, \quad U_0 = 1_G \tag{5.11}$$

for some left-invariant vector fields $X, Y_1, \dots, Y_m$ and independent Brownian motions $B^1, \dots, B^m$. Pick $f \in C^\infty(G)$. Then the Stratonovich differential of $f(U_t)$ is given by

$$df(U_t) = Df(dU_t) = DfX(U_t)dt + \sum_j DfY_j(U_t) \odot dB_t^j. \tag{5.12}$$

In order to avoid notational overload, we implicitly assume all involved functions defined on G to be depending on U_t, whenever this is oppurtune and simply write e.g. $f_t = f(U_t)$ instead. With the understanding that $Df(X(U_t)) = X_{U_t}(f)(= X(f)_t)$, we convert (5.12) into an Itô differential in order to simplify its treatment

$$df_t = X(f)_t + \sum_j Y_j(f)_t dB_t^j + \frac{1}{2}\sum_j d[Y_j(f), B^j]_t. \tag{5.13}$$

We extend f and the vector fields $X, Y_1, \dots, Y_n$ onto some open neighbourhood of $G \subset \mathbb{C}^{d\times d}$, and conduct our calculations in euclidean coordinates. Since U_t stays on G and $X, Y_1, \dots, Y_m$ are tangent to G, the specific choices for the extensions do not affect the following computations. We can express the vector fields Y_j in terms of the standard differentials ∂_i on $\mathbb{C}^{d\times d}$ via $Y_j = \sum Y_j^i \partial_i$ for some smooth $Y_j^i \in C^\infty(\mathbb{C}^{d\times d})$. With the aid of the identity Eq. (4.17) and the SDEs for the coordinate processes $(U_t^i)_t$ of $(U_t)_t$ we obtain from (5.11), we compute the summands of the drift correction seperately, exemplary for $Y = Y_1$ and $B = B^1$

$$\begin{aligned} d[Y(f), B]_t &= \sum_i \partial_i Y(f)_t d\left[U^i, B\right]_t = \sum_i \partial_i Y(f)_t d\left[\int dU^i, B\right]_t \\ &= \sum_i \partial_i Y(f)_t \left[\int X_s^i ds + \int \sum_l (Y_l^i)_s \odot dB_s^l, B\right]_t. \end{aligned}$$

We apply the Ito-Stratonovich conversion one more time, and end up with

$$\begin{aligned} d[Y(f), B]_t &= \sum_i \partial_i Y(f)_t \left[\int X_s^i ds + \sum_l \int \frac{1}{2} d[U^i, B^l]_s + \int \sum_l (Y_l^i)_s dB_s^l, B\right]_t \\ &\overset{*}{=} \sum_{i,l} \partial_i Y(f)_t d\int_0^t (Y_l^i)_s \underbrace{d[B^l, B]_s}_{=\delta_{1l} ds} = \sum_i (Y_1^i)_t \partial_i Y(f)_t dt = Y^2(f)_t dt. \end{aligned}$$

In the equality $(*)$, we use that the cross variations of the first two summands with the Brownian motion vanish, since these integrals have bounded first variation. Thus the Itô differential in (5.13)

is explicitly given by

$$df_t = \left(X(f)_t + \frac{1}{2} \sum_j Y_j^2(f)_t \right) dt + \sum_j Y_j(f)_t dB_t^j. \tag{5.14}$$

This allows us to compute the infinitesimal changes of the expectation of $f_t = f(U_t)$. Because the SDE (5.11) is left-invariant, the constant left shift $l_g(U_t) = gU_t$ satisifes the the same stochastic differential equation as U_t, only with the initial condition replaced by $U_0 = g$. Therefore we get

$$\begin{aligned}
\mathcal{L}f(g) &= \lim_{\delta \to 0} \frac{1}{\delta} \left(\mathbb{E}[f(gU_\delta)] - f(g) \right) = \lim_{\delta \to 0} \frac{1}{\delta} \left(f(g) \int_0^t df_s - f(g) \right) \\
&= \lim_{\delta \to 0} \frac{1}{\delta} \mathbb{E} \left[\int_0^\delta (Xf)(gU_s) + \frac{1}{2} \sum_j (Yf)(gU_s) ds + \int_0^\delta \sum_j (Y_j f)(gU_s) dB_s^j \right] \\
&= \lim_{\delta \to 0} \mathbb{E} \left[\frac{1}{\delta} \int_0^\delta (Xf)(gU_s) + \frac{1}{2} \sum_j \left(Y_j^2 f \right)(gU_s) ds \right],
\end{aligned}$$

where we use that the expectation of the Itô integral vanishes in the last step. By dominated convergence, we can take the limit inside the expectation and are left with

$$\mathcal{L}f = \frac{d}{dt}\Big|_{t=0} P_t f = Xf + \frac{1}{2} \sum_j Y_j^2 f. \tag{5.15}$$

We have just proven the following proposition.

Proposition 5.6. For the left Lévy process $(U_t)_t$ defined as the solution to the Stratanovich SDE (5.11), the infinitesimal generator is given by the left-invariant differential operator

$$\mathcal{L} = X + \frac{1}{2} \sum_j Y_j^2. \tag{5.16}$$

Remark 5.7. Since the assignment $x \mapsto X \in Der(C^\infty(G))$ essentially defines a faithful representation of the Lie algebra $\mathfrak{g}$ on $C^\infty(G)$, we can view $\mathcal{L}$ in (5.16) as the action of the element

$$\mathfrak{k} = x + \frac{1}{2} \sum_j y_j^2 \tag{5.17}$$

in the universal envelopping algebra $U(\mathfrak{g})$ under this representation. Given an ad-invariant inner product $\langle \cdot, \cdot \rangle$ on $\mathfrak{g}$, we can characterise Brownian motion on G as processes generated by the associated Laplacian Δ of the corresponding metric on G. The Laplacian Δ is just the action of the Casimir element $C_\mathfrak{g}$ with respect to $\langle \cdot, \cdot \rangle$ under the above representation. Because the $C_\mathfrak{g}$ lies in the center of $U(\mathfrak{g})$, Δ commutes with all left-invariant differential operators on G, which

expresses the highly symmetrical nature of Brownian motion.

Because the diffusion matrix A in Thm. 5.4 is symmetric, we can always find a basis in which it is diagonal so that the generator of a continuous Lévy process takes the form (5.16). In particular, with the converse statement of Thm. 5.4 this shows that every continuous Lévy processes can be modelled as a weak solution to a Stratonovich SDE as in (5.11). The generator of a continuous Lévy process is a second-order differential operator with dense domain of definition $\mathrm{Dom}(\mathcal{L}) \subset L^2(G)$. It is well known that such differential operators are *closeable* i.e. the closure of its graph in $L^2(G) \oplus L^2(G)$ is the graph of some operator with larger domain (c.f. p. 78 in [66]). This condition ensures the existence of a densely defined L^2-adjoint $\mathcal{L}^\dagger$ (Chapter 13 in [62]) so that $\int_G \mathcal{L}(f_1) f_2 d\mu_G = \int_G f_1 \mathcal{L}^\dagger(f_2) d\mu_G$ holds for all $f_1 \in \mathrm{Dom}(\mathcal{L})$, $f_2 \in \mathrm{Dom}(\mathcal{L}^\dagger)$. In order to explicitly determine the action of $\mathcal{L}^\dagger$, we observe that the left-invariant vector fields $X = x^l$ inside $L^2(G)$ are formally skew-adjoint.[1] That is we have

$$\langle f, X(g) \rangle_{L^2} = -\langle X(f), g \rangle_{L^2} \text{ for all } f, g \in C^\infty(G),$$

one of which we assume to have compact support. Consequently the Y_j^2 are formally self adjoint and we arrive at

$$\mathcal{L}^\dagger = -X + \frac{1}{2} \sum_j Y_j^2.$$

In particular we see that $C_c^\infty \subset \mathrm{Dom}(\mathcal{L}^\dagger) \subset \mathcal{L}^2(G)$ and that the second order part of the generator $\mathcal{L}$ is always negative semi-definite. What do we gain by studying the generator of a transition semigroup? In the framework of Markov processes it is a central task to analyse the long-terms dynamics of a process, the existence of stationary distributions as well as the convergence to those distributions. It turns out that in the time-homogenous case, these questions entirely traced back to the generator $\mathcal{L}$. Specifically, the forward equation (5.9) allows to infinitesimally characterise the time-evolution of the probability distribution μ_t itself, as we explain in the following. Since probabibility measures are distributions in a functional analytic sense, it indeed makes sense to apply differential operators to probability measures. We begin with the observation that the mapping $t \mapsto \mu_t$, $\mathbb{R}_{\geq 0} \to \mathcal{M}^1(G) \subset \mathcal{D}'(G)$ is continuous in the weak topology, i.e.

$$\lim_{t \to 0} \langle P_t^* \mu_0 - \mu_0, f \rangle = \lim_{t \to 0} \langle \mu_0, P_t f - f \rangle = 0$$

for every $f \in C_0(G)$. A priori it is not clear if or why it also should be strongly continuous or

[1] For those familiar with geometry, let ω be the left-invariant n-form corresponding to the Haar measure and $\pounds$ the Lie derivative. Since both X and ω are left invariant, $\pounds_X \omega = 0$, and

$$X(h)\omega = \pounds_X(h\omega) - h\pounds_X\omega = d\iota_x h\omega + \iota_X dh\omega = d\iota_X h\omega.$$

If we now choose a submanifold with boundary N which contains the support of $X(fg)$, we get

$$\int_G (X(f)g + fX(g)) d\mu_g = \int_M X(fg)\omega_G = \int_{\partial G} fg\iota_X\omega = 0$$

by Stoke's Theorem.

even differentiable. Therefore, we consider the pairing (which is realised by integration) with a fixed, rapidly decaying function $f \in C^\infty(G)$ first. According to the forward equation (5.9) we get

$$\partial_t \langle \mu_t, f \rangle = \partial_t P_t f = P_t \mathcal{L} f = \langle \mu_t, \mathcal{L} f \rangle,$$

and with another iteration

$$\partial_t^2 \langle \mu_t, f \rangle = \langle \mu_t, \mathcal{L}^2 f \rangle,$$

so the map is twice weakly continuously differentiable. With the aid of the result on weak-to-strong differentiability (see Lemma B.3) we can conclude that the map $t \mapsto \mu_t$ is *strongly* differentiable, and therefore

$$\langle \partial_t \mu_t, f \rangle = \partial_t \langle \mu_t, f \rangle = \langle \mathcal{L}^\dagger \mu_t, f \rangle.$$

As this holds for all rapidly decaying functions $f \in C^\infty(G)$, we can conclude that the map $t \mapsto \mu_t$ satisfies the differential equation

$$\partial_t \mu_t = \mathcal{L}^\dagger \mu_t. \tag{5.18}$$

It is clear that a measure ν is a stationary distribution of the process if and only if $\mathcal{L}^\dagger \nu = 0$. If the distribution μ_t of the Levy process $(g_t)_t$ admits a density p_t with respect to the invariant Haar measure μ_G, we are therefore left with the partial differential equation

$$\partial_t p_t = \mathcal{L}^\dagger p_t. \tag{5.19}$$

This version of the forward equation sometimes appears under the name *Fokker-Planck equation* in the physics literature.

Conclusion

We have seen that Lévy processes on Lie groups give rise to strongly continuous probability operator semigroups which are completely characterised by their left-invariant infinitesimal generator $\mathcal{L}$. Proposition 5.6 clarifies how these generators look for the processes we are interested in. Most importantly, the generator determines the local spread of the induced probability distribution on the state space via the Fokker-Planck equation. Therefore it plays an important role in the analysis of the convergence to equilibrium.

Part III

CONVERGENCE OF LEFT-INVARIANT DIFFUSION PROCESSES ON COMPACT LIE GROUPS

In this part we analyse under which assumptions we can conclude convergence of continuous Lévy processes to the Haar measure on compact Lie groups. We begin our discussion in Chapter 6 by analysing total variation and Wasserstein distances, which are standard analytical tools to characterise convergence. The approach of Chapter 7 is more algebraic, where we discusses convergence of the moments in view of Chapter 3, building on [19]. In turns out that we need the *same* condition in both cases, specifically the regularity of the control system in the sense of Def. 2.8, which can be exploited in very different manners.

6. CONVERGENCE IN TOTAL VARIATION DISTANCE

Diffusion processes on manifolds are continuous time-homogenous Markov processes which describe, how an initial probability distribution spreads over the state space under the influence of a second-order differential operator with negative semidefinite second order part. The first order component defines a non-random flow on the manifold, whereas the second-order is responsible for the 'stochasticness' and the spreading of probability. The driving differential operator is just the infinitesimal generator $\mathcal{L}$ of the process which we introduced in the previous chapter. Continuous Lévy processes on Lie groups are the special case of such diffusions where the generator $\mathcal{L}$ is *left-invariant* and the Markov kernels act by convolutions. The central problem in the theory of diffusion processes is to determine if and how fast a diffusion converges to its stationary distribution, provided it exists. It is clear that the Haar measure on a compact group is invariant under convolutions and therefore stationary. If on the other hand the support of a Lévy process is a non-compact Lie group G, it is not difficult to show that such an invariant distribution ceases to exist. From now on we consider Lévy processes on compact Lie groups which start at the identity 1_G, unless explicitly indicated differently.

6.1 Hypoellipticity

The long time behaviour of diffusion processes is strongly linked to functional analytic properties of its infinitesimal generator. In this context, different notions of ellipticity, in particular hypoellipticity play an important role. Recall that a differential operator P acting on distributions $u \in \mathcal{D}'$ is called *hypoelliptic*, if it holds that Pu is smooth only if u is alread given by a smooth function. The central tool in order to determine hypoellipticity of a second-order differential operator is the celebrated *Hörmander condition*.

Theorem 6.1 (Hörmander 1967, [67])**.** Let $(P, \mathrm{Dom}(P))$ be a second order differential operator on a Riemannian manifold (M, q), which can be written as

$$P = c + W + \sum_i V_i^2$$

for some vector fields $W, V_1, \ldots, V_n$ and $c \in C^\infty(M)$. *Then P is hypoelliptic if and only if $W, V_1 \ldots, V_n$ satisfy the Hörmander condition i.e. the Lie closure of the vector fields $W, V_1, \ldots, V_n$ pointwise spans the tangent space of M.*

For a Lévy process starting at a single point, it is always possible to achieve hypoellipticity of its generator by considering the smallest submanifold on which the process is supported. In particular, this can be achieved for the process $(U_t)_t$ defined earlier (or again in (6.1) if we restrict ourselves to the subgroup $G_\Sigma = G$ that can be controlled by the system $(x|y_1, \dots, y_m)$.

Remark 6.2. In several situations one demands for a slightly stronger property. For vector fields $W, V_1, \dots, V_n$ on M consider

$$\mathcal{V}_0 := \{V_i : i = 1, \dots, n\} \text{ and inductively } \quad \mathcal{V}_{j+1} := \{[Z, \mathcal{V}_j], Z \in \{W, V_1, \dots, V_n\}\}\,.$$

The vector fields $(W|V_1, \dots, V_n)$ satisfy the *parabolic Hörmander condition* if

$$\mathcal{I}_0 := \bigcup_{j \geq 0} \mathcal{V}_j$$

spans T_pM for every $p \in M$. In the language of control systems, $\mathcal{I}_0$ is just the general appearance of the *zero-time ideal* $\mathfrak{I}_0$ we defined for left invariant systems (Def. 2.4).

With this in mind, we introduce the following distinctions.

Definition 6.3. Let $(g_t)_t$ be a Lévy process on a Lie group G with generator $\mathcal{L}$. We say that $(g_t)_t$ is

1. *non-degenerate/elliptic* if $\mathcal{L}$ is elliptic, i.e. the matrix A in Thm. 5.10 is strictly positive definite.
2. *strongly hypoelliptic* if the diffusive part $\sum_i Y_i^2$ of $\mathcal{L}$ is hypoelliptic.
3. *weakly hypoelliptic* if the operator $\mathcal{L} - \partial_t$ is hypoelliptic on $G \times \mathbb{R}_{>0}$.

It is clear that

$$\text{elliptic} \implies \text{strongly hypoelliptic} \implies \text{weakly hypoelliptic}$$

and the inclusions are generally strict.

Remark 6.4. Hypoelliptic Lévy processes in the sense of what we call strongly hypoelliptic, are investigated for example in [59]. Translated into the language of control theory this assumption essentially means, that the underlying control system is assumed to be strongly controllable. From the point of view of a physicists (or quantum engineer for that sake), the requirement of strong controllability is ultimately unsatisfactory. In applications, the drift is normally used to couple different subsystems, and therefore essential to generate the dynamical Lie algebra. Fortunately, strong hypoellipticity of $(g_t)_t$ is slightly too restrictive to analyse convergence properties in general. In the following we show that weak hypoellipticity is the most general case under which convergence to equilibrium in total variation is guaranteed. In Chapter 7, we show that this is also necessary by providing abstract counterexamples.

We just discussed that strongly hypoelliptic Lévy processes arise in strongly controllable systems. What is the condition we need to impose on the system to arrive at a weakly hypoelliptic diffusion?

Proposition 6.5. For a left-invariant controllable system $(X|Y_1 \dots, Y_m)$ on a Lie group G, the Lévy process subject to the stochastic differential equation

$$dU_t = X(U_t)dt + \sum_j Y_j(U_t) \odot B_t^j, \quad U_0 = 1_G \tag{6.1}$$

with independent Brownian motions $B^1, \dots, B^m$, is weakly hypoelliptic if and only if the zero-time ideal satisfies $\mathfrak{I}_0 = \mathfrak{g}$, i.e. the control system is regular in the sense of Def. 2.8.

Proof. The Lie algebra of $G \times \mathbb{R}$ is given by the trivial extension $Lie(G \times \mathbb{R}) = \mathfrak{g} \oplus \mathbb{R}$ where $\oplus$ denotes the sum of vector spaces. Let τ denote a generator of the $\mathbb{R}$ component, such that ∂_t is the corresponding invariant vector field. By Hörmanders Theorem 6.1 the operator

$$\mathcal{L} - \partial_t = X - \partial_t + \sum_j Y_j^2$$

is hypoelliptic if and only if $\mathfrak{h} := \langle \tau + x, y_1, \dots, y_m \rangle_{Lie} = \mathfrak{g} \oplus \mathbb{R}$. Since τ commutes with all the other generators, we have immediately

$$\mathfrak{h} = \mathfrak{I}_0 \oplus \mathrm{span}_{\mathbb{R}}(\tau + x)$$

If we now assume the control system to be regular, we have $\mathfrak{I}_0 = \mathfrak{g}$ and hence

$$\mathfrak{h} = \mathfrak{g} \oplus \mathrm{span}_{\mathbb{R}}(\tau + x) = Lie(G \times \mathbb{R}).$$

Conversely, if the system is not regular, $\mathfrak{I}_0 \subset \mathfrak{g}$ has codimension one, therefore $\mathfrak{h} \subsetneq \mathfrak{g} \oplus \mathbb{R}$ and $\mathcal{L} - \partial_t$ cannot be hypoelliptic. □

As we discussed in Remark 2.9, every controllable system on a semisimple Lie group is also regular. In particular, on semisimple Lie groups we have that $\mathcal{L} - \partial_t$ is hypoellipctic if and only if $\mathcal{L}$ is hypoelliptic.

Why is this type of hypoellipticity benefitial for us? As we have seen already, the adjoint of the generator $\mathcal{L}$ with respect to the $L^2(G)$ inner product is given by simply changing the sign of the degree one part. In particular, $\mathcal{L} - \partial_t$ is hypoelliptic if and only if $\mathcal{L}^\dagger - \partial_t$ is hypoelliptic. The law μ_t of a diffusion process obeys the Fokker-Planck equation $\mathcal{L}^\dagger \mu_t = \partial_t \mu_t$ which we derived in end of Chapter 5. If now $\mathcal{L}^\dagger - \partial_t$ is hypoelliptic, μ_t has empty singular support for all $t > 0$ and therefore admits a smooth density with respect to the Haar measure.

Corollary 6.6. The law μ_t of a Lévy process $(U_t)_t$ on a Lie group G induced by a regular stochastic control system via the SDE (6.1), admits a smooth density p_t with respect to the Haar measure for all times $t > 0$.

Despite the fact that we derived this result borrowing from the theory of partial differential equations, the result is entirely probabalistic in nature. Today's standard proof of Hörmanders Theorem is conducted via the stocastic variational calculus which is also known as *Malliavin calculus*, [68, 69]. In this sense, the smoothness of the density is substantially a consequence of the following theorem.

Theorem 6.7 (Theorem 3.2 in [69])**.** Let $(W|V_1, \dots, V_m)$ be a control system on $\mathbb{R}^d$ such that the smooth vector fields $W, V_1, \dots, V_m$ satisfy the parabolic Hörmander condition in Rem. 6.2. Then any solution $(Z_t)_t$ to the Stratonovich SDE

$$dZ_t = W(Z_t)dt + \sum_j V_j(Z_t) \odot dB_t^j,\ Z_0 = z_0 \tag{6.2}$$

admits a smooth density with respect to the Lebesgue measure on $\mathbb{R}^d$ for all times $t > 0$.

Since the statement is local in nature and Stratonovich integration is compatible with smooth coordinate transformations, the result transfers to smooth manifolds and in particular Lie groups analogously.

Remark 6.8. Consider again an invariant control system Σ on a compact Lie group, which we assume to be regular or equivalently that the vector fields satisfy the parabolic Hörmander condition. The support Theorem 4.22 implies that at time $t > 0$ the support of the law μ_t of the solution to (6.1) is given by the attainable set $\mathcal{A}_t$. Following Prop. 2.6 we can conclude that there exists a time $T > 0$ such that the support of μ_t is equal to G for all $t \geq T$. Therefore the density p_t of μ_t not only exists for all times $t > 0$, but we can also find a time $T > 0$ such that p_T vanishes only on a set of measure zero. This is already sufficient to conclude that $\inf_{g \in G} p_(g)t > 0$ holds for all $t > T$. Exploiting the fact that the law of a Lévy process forms a continuous convolution semigroup, such lower bounds on the density provide a minimum spread of the probability distribution. This can be regarded as the driving force behind convergence to the equilibrium in our setting, which we formally use in Prop. 6.11.

Deriving lower bounds on the density after a sufficiently long time actually does not require smoothness or even continuity of the density. In fact, already the *existence* of a density is sufficient for this.

Lemma 6.9 (Strongly inspired by Prop. 23 in [70])**.** Assume that a measure ν on a compact Lie group G admits a density $\varrho \in L^2(G)$. Then we find a number $n \in \mathbb{N}$ and constant $c > 0$ such that the lower bound $\varrho^{*n}(g) \geq c$ on the density of ν^{*n} holds for almost all $g \in G$.

Proof. In the following, all estimates should be interpreted to hold up to a set of measure zero. We fix a bi-invariant metric $d(\cdot,\cdot)$ on G. Let $t>0$ and choose $g_0 \in G$ such that the L^2-density of ν satisfies $\varrho \geq c > 0$ on an ε-ball around g_0. The properties of the metric $d(\cdot,\cdot)$ enable us to derive a lower bound of the density ϱ^{*2} of $\mu_t * \mu_t$ on the ball of radius $3r/2$ around g_0^2. For $h \in B_{3r/2}(g_0^2)$ we can estimate

$$\varrho^{*2}(h) = \int \varrho(g)\varrho(hg^{-1})dg \geq \alpha^2 \mu_G\left(\{g \in B_r(g_0) : hg^{-1} \in B_r(g_0)\}\right).$$

Observe that $hg^{-1} \in B_r(g_0)$ if and only if $g \in B_r(g_0^{-1}h)$. We therefore only need to show that the Haar measure of the intersection $B_r(g_0) \cap B_r(g_0^{-1}h)$ is strictly positive. Because the exponential flow corresponds to geodesic curves, we find $x \in \mathfrak{g}$ with $\|x\| \leq \frac{3}{2}r$ such that $h = g_0^2 \exp(x)$. We now show that

$$B_{r/4}(g_0 \exp(x/2)) \subset B_r(g_0) \cap B_r(g_0^{-1}h). \tag{6.3}$$

Lett $g \in B_{r/4}(g_0 \exp(x/2))$. Then we get the estimates

$$d(g,g_0) \leq d(g, g_0\exp(x/2)) + d(g_0\exp(x/2), g_0) \leq \frac{r}{4} + 3\frac{r}{4} = r,$$

and since $h = g_0^2\exp(x)$ also

$$\begin{aligned} d(g, g_0^{-1}h) &\leq d(g, g_0\exp(x/2)) + d(g_0\exp(x/2), g_0\exp(x)) \\ &\leq \frac{r}{4} + \|\frac{x}{2}\| \leq r, \end{aligned}$$

so (6.3) follows. Alltogether, for $h \in B_{3r/2}(g_0^2)$, we obtain the lower bound

$$\varrho^{*2}(h) \geq \alpha^2 \mu\left(B_{r/4}(g_0^{-1}h)\right) > 0.$$

Iteratively, we can proceed until $3r/2 \geq \operatorname{diam}(G)$ and we obtain a global bound after another iteration. □

Contrary to the preceeding discussion, this Lemma is applicable only to Lévy processes. This is simply due to the fact, that the transition kernels of general diffusions do not necessarily act by convolutions.

6.2 Convergence via Relative Entropy

We proceed by showing that as soon a distribution ν on a compact group admits a μ_G-density, the convolution sequence ν^{*n} converges exponentially to the Haar measure in total variation. The key idea is to estimate the total variation distance of ν and μ_G against the relative entropy of ν with respect to μ_G and exploit monotonicity of the latter under convolutions. Let us briefly recall the

definition of the relative entropy:

Definition 6.10. Let μ, ν be probability measures on some measureable space $(S, \mathscr{A})$. If μ is absolutely continuous with respect to ν ($\mu \ll \nu$), we define the *Kullback Leibler divergence*, or *relative entropy* to be

$$\mathcal{D}_{KL}(\mu||\nu) := \int_S \log\left(\frac{d\mu}{d\nu}\right) d\mu. \tag{6.4}$$

In case such a density does not exist, we set $\mathcal{D}_{KL}(\mu\|\nu) := \infty$.

The relative entropy has several useful properties such as convexity and monotonicity under the application of probability kernels, and provides bounds on the total variation distance. A summary of these results can be found in the Appendix, Lemma B.14, Prop B.15 and Prop B.16. For the special case of convolution kernels monotonicity means that

$$\mathcal{D}_{KL}(\mu * \eta||\nu * \eta) \leq \mathcal{D}_{KL}(\mu||\nu) \tag{6.5}$$

holds for all $\mu, \nu, \eta \in \mathcal{M}^1(G)$. The key observation to conclude convergence is, that this inequality actually needs to be strict, if we have a lower bound on the density of η with respect to the Haar measure.

Proposition 6.11 (See [70], Lemma 22)**.** Let ν be a probability measure on a compact Lie group G, which is absolutely continuous with respect to the Haar measure μ_G. Assume that we can minorise the density of ν by $\frac{d\nu}{d\mu_G} \geq c$ for some positive constant c, up to a set of measure zero. Then there exist constants $C, \alpha > 0$ such that we can estimate the total variation distance of $\nu_n := \nu^{*n}$ to the Haar measure by

$$d_{TV}(\mu_G, \nu_n) \leq Ce^{-\alpha n}, \tag{6.6}$$

and therefore the convolution sequence ν_n converges to the Haar measure in total variation distance.

Proof. With the aid of Pinsker's inequality (see Proposition B.16), we estimate

$$d_{TV}(\nu_n, \mu_G) \leq \sqrt{\frac{1}{2}\mathcal{D}_{KL}(\nu_n||\mu_G)}.$$

Next we consider $\eta := (\nu - c\mu_G)(1-c)^{-1}$ which defines a probability measure by our minorisation assumption. This allows to write ν as a convex combination of measures $\nu = c\mu_G + (1-c)\eta$. Using the convexity of $\mathcal{D}_{KL}$ (Lemma. B.14) we get for every n:

$$\begin{aligned}
\mathcal{D}_{KL}(\nu * \nu_n||\mu_G) &\leq c\mathcal{D}_{KL}(\mu_G * \nu_n||\mu_G) + (1-c)\mathcal{D}_{KL}(\eta * \nu_n||\mu_G) \\
&= c\mathcal{D}_{KL}(\mu_G\|\mu_G) + (1-c)\mathcal{D}_{KL}(\eta * \nu_n\|\eta * \mu_G) \\
&\leq (1-c)\mathcal{D}_{KL}(\nu_n\|\mu_G),
\end{aligned}$$

where we use monotonicity of the entropy (Prop. B.15) in the last step. Inductively, we arrive at $\mathcal{D}_{KL}(\nu_{n+1}\|\mu_G) \leq (1-c)^n \mathcal{D}_{KL}(\nu\|\mu_G)$, and therefore

$$d_{TV}(\nu_{n+1}, \mu_G) \leq (1-c)^{n/2} \sqrt{\frac{1}{2}\mathcal{D}_{KL}(\nu\|\mu_G)} \tag{6.7}$$

for all n. Clearly (6.6) holds with $C := (\mathcal{D}_{KL}(\nu\|\|\mu_G)/2)^{1/2}$ and $\alpha := \log(1-c)/2$. □

Combining this proposition with Lemma 6.9, we arrive at the following theorem.

Theorem 6.12. For a weakly hypoelliptic Lévy process $(g_t)_{t\geq 0}$ on a compact Lie group G, the associated convolution semigroup $\{\mu_t\}_t$ of its law converges exponentially to the Haar measure in total variation, i.e. we find contants $C, \alpha > 0$ with

$$d_{TV}(\mu_t, \mu_G) \leq Ce^{-\alpha t} \to 0 \text{ as } t \to \infty. \tag{6.8}$$

Since G is compact, we can estimate $\mathcal{W}_1 \lesssim d_{TV}$, (see remark after Prop. B.11) therefore (6.8) implies exponential convergence in the L^1-Wasserstein distance as well.

Proof. Because $(g_t)_{t\geq 0}$ is weakly hypoelliptic, the density of $\mu_t = Law(g_t)$ exists for all times $t > 0$. With the semigroup property $\mu_{t+s} = \mu_t * \mu_s$ and Lemma 6.9, we find a time $T > 0$ such that the density of μ_T meets the requirements of Proposition 6.11. The relative entropy $\mathcal{D}_{KL}(\mu_t||\mu_G)$ is monotonically decreasing in t, and therefore the exponential convergence of the sequence $(\mu_{nT})_n$ carries over to the complete semigroup $\{\mu_t\}_{t\geq 0}$. □

Although this result looks nice, explicitly determining the convergence rate is rather challenging, even though we are working with Lie groups which are comparatively rich in structure.

6.3 Some Remarks on Convergence Rates

Since the following only contains an overview of unsuccessful approaches, we will not explicitly state all definitions and results mentioned and just give references where these can be found. The aim of this section is to show that despite several available methods for deriving bounds on the convergence of hypoelliptic diffusion proceses, none of them actually quite fits the setting we are interested in.

Non-degenerate case

If the considered process is non-degenerate it is not difficult to obtain bounds on the convergence. In this case, Bakry and Eméry [25] devised a method determining explicit convergence rates using log-Sobolev inequalities. The central objects in their analysis are the *square-field operator*

Γ ('carree du champ'), which serves as a generalisation of the Dirichlet form $q(\nabla f, \nabla g)$ and is defined by

$$\Gamma(f,g) := \mathcal{L}(fg) - f\mathcal{L}(g) - g\mathcal{L}(f), \tag{6.9}$$

and its iteration

$$\Gamma_2(f,g) := \mathcal{L}(\Gamma(f,g)) - f\mathcal{L}(g) - g\mathcal{L}(f). \tag{6.10}$$

With these tools, they managed to show that the transition operators P_t of the diffusion associated to $\mathcal{L}$ are L^2-Wasserstein contractive, i.e.

$$\mathcal{W}_2(P_t^*\mu, P_t^*\nu) \leq e^{-Kt}\mathcal{W}_2(\mu,\nu) \text{ for all } \mu,\nu \in \mathcal{M}^1(M). \tag{6.11}$$

holds if and only if for some $K > 0$ the gradient bound $\Gamma_2(f,f) \geq K\,\Gamma(f,f)$ is satisfied for all smooth f. For the special case of Brownian motion, $\mathcal{L} = \Delta$ is just the Laplacian associated to the Riemannian metric q. Here, as shown for instance in [71], the gradient bound is equivalent to a lower bound on the Ricci curvature:

$$\mathrm{Ric}_q \geq Kq.$$

In particular, on a compact Lie group G with a bi-invariant metric we can relate the Ricci curvature to the Killing form $\mathrm{Ric}_q = -\frac{1}{4}\mathcal{K}$ (cf. [72]) and therefore find such a constant $K > 0$ without difficulties if G is semisimple.

General gradient estimates

The problem with the methods above, is that they generally require ellipticity of $\mathcal{L}$ in one way or another as noted in [21], which does not apply to the case we are ultimately interested in. Slightly altering the definition of the square field operator, Baudoin [21] managed to provide conditions that can be applied to properly hypoelliptic diffusion operators $\mathcal{L}$. Instead of the square field operator, he considered the bilinear form $T(f,g) := \sum a_{ij}\partial_i f \partial_j g$ for some positive definite matrix A and its iteration with respect to the generator $\mathcal{L}$ which is given by

$$T_2(f) := \mathcal{L}(T(f,f)) - T(f,\mathcal{L}f). \tag{6.12}$$

Together with some technical assumptions, the gradient bound $T_2(f) \geq KT(f,f)$ then leads to the same Wasserstein-contraction properties as before

$$\mathcal{W}_2(P_t^*\mu, P_t^*\nu) \leq e^{-Kt}\mathcal{W}_2(\mu,\nu). \tag{6.13}$$

Checking the condition $T_2(f) \geq K\,T(f,f)$ is not exactly trivial. Several approaches to derive such estimates are discussed in [21, 22]. We briefly sketch the ones which appeared to be the most promising for our scenario.

Attempt 1:

For the Haar measure on a semisimple compact Lie group, it is not difficult to derive a Poincaré inequality

$$\int f^2 d\mu_G - \left(\int f d\mu_G \right)^2 \leq C \int |\nabla f|^2 d\mu_G, \tag{6.14}$$

for some constant $C > 0$ which again depends on the curvature, see for instance [73]. As argued in [21], the gradient bound $T_2(f) \geq KT(f, f)$ can then be relaxed to a similar estimate under integration

$$\int T_2(f) d\mu_G \geq K_1 \int T(f, f) d\mu_G - K_2 \int \Gamma(f) d\mu_G, \tag{6.15}$$

where Γ again denotes the square-field operator of $\mathcal{L}$. With this assumption (6.15), [21] shows the convergence of the density $p_t \to 1$ in L^2 with an exponential rate $K := 2 \min\{K_1, K_2/C\}$. Again it is not clear how to check (6.15). Even in the simplest example where we choose $G = \mathrm{SU}(2)$ with cyclic generators $x_1, x_2, x_3 \in \mathfrak{su}_2$, and $\mathcal{L} = X_1 + X_2^2$ neither of canonical choices $T(f, g) = q(\nabla f, \nabla g)$, or $T(f, g) = \sum_j X_j(f) X_j(g)$ has produced any results of value.

Attempt 2:

If the system is strongly controllable, the diffusive part of the generator $\mathcal{L}$ is the sub-Laplacian associated to the sub-Riemannian structure that arises in strongly controlled systems. In order to analyse the diffusion semigroups generated by such sub-Laplacians, Baudoin et al [23] introduced so called *generalised curavture dimension inequalities*. These are in general not easy to derive, but for instance if $G = SU(2)$ and $\mathcal{L} = X^2 + Y^2$ ($x, y, z \in \mathfrak{su}_2$ cyclic generators as usual) their methods in [21, 23] ultimately yield L^2-Wasserstein contractivity

$$\mathcal{W}_2(P_t^* \mu, P_t^* \nu) \leq e^{-t/2} \mathcal{W}_2(\mu, \nu) \tag{6.16}$$

with exponential rate $\frac{1}{2}$.

Attempt 3:

One can also apply the sub-Riemannian setting in a different fashion. Specifically, we assume that the controls form a basis for the control algebra $\mathfrak{g}_0$ (e.g the Examples A, D, E of our simulations in the introduction). In this situation, the controlled distribution $\mathcal{Y} \subset TG$ is closed under the commutator and has the same dimension everywhere. Then it follows from Frobenius Theorem[1] that $\mathcal{Y}$ determines a *foliation*, that is a disjoint collection of submanifolds $\{G_i\}_{i \in I}$ (here copies of the subgroup generated by the controls) such that $TG_i = \mathcal{Y}|_{G_i}$. Geometrically this amounts to having a fiber bundle $\pi : G \to G/G_0 =: M$. We proceed by assuming that the orthogonal complement $\mathcal{H} \subset TG$ contains the drift X and that $\mathcal{H}$ is *bracket-generating*, that is $Lie(\mathcal{H})(g) =$

[1] see for instance [74], Theorem 19.12

T_gG for all $g \in G$. Because geodesics on G have the form $t \mapsto g\exp(tz)$ for some $z \in \mathfrak{g}$, any geodesic tangent to a fiber will completely stay in this fiber, thus the foliation $\{G_i\}_i$ is *totally geodesic*, and fits into the setting of [22]. We can write the diffusion operator $\mathcal{L}$ as

$$\mathcal{L} = \Delta_{\mathcal{Y}} + X,$$

where $\Delta_{\mathcal{Y}}$ is the *vertical* Laplacian on the G_i is the (horizontal) drift X. If $\nabla_{\mathcal{H}}$, $\nabla_{\mathcal{Y}}$ denote the projections of the gradient ∇ onto the respective subspaces, following the computations of [22] we can write

$$T_2(f) = |\nabla_{\mathcal{Y}}\nabla_{\mathcal{H}}f|^2 + |\nabla^2_{\mathcal{Y}}f|^2 + \mathrm{Ric}_{\mathcal{Y}}(\nabla f, \nabla f) + DX(\nabla f, \nabla f). \tag{6.17}$$

The last term $DX(V,W) := q\left(\nabla^{LC}_V X, W\right)$, which involves the Levi-Civita connection ∇^{LC}, *vanishes* because X is left-invariant (follows from Prop.2.26 in [72]). The Ricci-curvature $\mathrm{Ric}_{\mathcal{Y}}$ is zero as well, because the leaves are abelian Lie groups. In particular Thm. 7.3. in [22] is not helpful in this case, and in order to derive estimates $T_2(f) \gtrsim |\nabla f|^2$, we would need bounds of the type $|\nabla_{\mathcal{Y}}\nabla_{\mathcal{H}}f| \gtrsim |\nabla_{\mathcal{H}}f|$. This cannot be expected, as one can infer from functions where the first-order horizontal derivatives are constant on the leaves of the foliation at least on some open neighbourhood. Functions exhibiting these properties can be constructed in local coordinates.

Another remark - Ergodicity

Remark 6.13. While for many applications it may suffice to simulate the uniform distribution using *ergodicity* of the Markov process, this is not the case for randomised benchmarking. Recall that a continuous time markov process $X_s : \Omega \to S$ on a measureable state space S with invariant measure μ is ergodic if it satisfies the following law of large numbers

$$\frac{1}{T}\int_0^T f(X_s)ds \to \mathbb{E}_\mu[f] \text{ as } T \to \infty \tag{6.18}$$

for all measureable $f : S \to \mathbb{R}$ and every initial distribution of X_0. A sufficient criterion to determine if a Markov process is ergodic is that its generator $\mathcal{L}$ is hypoelliptic (Prop. 3.1.13 in [20]), which is certainly the case for the situations we are interested in. The problem we encounter when trying to use this (or rather a discrete version of (6.18)) in randomised benchmarking protocols such as in Tab. 1 is the following: When we measure the survival probability of the initial state under the noisy implementation and inversion of the gate U_s for some $s > 0$ (which amounts to determining $f(X_s)$ in (6.18)), the system is expected to collapse back to the ground state in most of the cases. In particular, the effective evolution starts again at the identity 1_G and not at U_s, therefore (6.18) is not applicable.

7. CONVERGENCE OF THE MOMENT SEMIGROUP

This section continues the discussion of diffusion processes on compact Lie groups, where we combine the results of Chapter 5 with the framework introduced in Chapter 3. Specifically we consider the moments of such diffusions in order to show that they provide a natural source for exponentially converging approximate k-Haar measures. The situation will be similar to the one we found before. The assumption of weak hypoellipticity allows us to prove exponential convergence to the Haar measure in all moments. Unless the diffusion is non-degenerate, it is generally hard to computationally determine the convergence rate, even though it is much clearer what quantities we need to look at compared to convergence in total variation/Wasserstein distances. A detailed discussion of the non-degenerate case is presented in [18], where the authors prove that for mixing times of order $\mathcal{O}(poly(k), \log(d), \log(1/\varepsilon))$, a Brownian motion (with an optional drift) produces an ε-approximate k-Haar measure on $\mathrm{SU}(d)$, see Theorem 9 therein. Their analysis of the convergence rate relies fundamentally on the fact that for non-degenerate diffusions, the infinitesimal generator $\mathcal{L}$ can be linked to the Casimir element of $\mathfrak{su}_d$ in analogy to Remark 5.7. Exploiting the distinct role the Casimir element plays in the representation theory of Lie algebras one can then determine the convergence rates under different representations. Unfortunately this approach cannot be adapted to the degenerate case, leaving us with a greater applicability but overall weaker results.

The situation with a degenerate control system on $\mathrm{SU}(d)$ is already discussed in [19]. Specifically, the authors analyse probability distributions on $\mathrm{SU}(d)$, generated by applying stationary (which is called 'harmonic' in [19]) and normally distributed stochastic processes c_j independently to the controls of the system, resulting in the random evolution of a mixed state $\rho(t)$

$$\frac{d}{dt}\rho(t) = -i[H, \rho(t)] - i\sum_{j=1}^{n}[V_j, \rho(t)]c_j(t). \tag{7.1}$$

The processes c_j can for instance be modelled by Ornstein-Uhlenbeck-processes[1] which have undergone a sufficiently long mixing time. Subsequently [19] presents an outline to show that the qulitative result on exponential convergence of [18] still holds in this setting. In some point of their computations, the authors assume that the autocorrelation of the noise terms is close to the Dirac delta distribution. This is one of the chracteristic features of white noise, therefore it is

[1] Ornstein-Uhlenbeck processes are solutions to Itô-SDEs of the type $dZ_t = \theta(\alpha - Z_t) + \beta dB_t, Z_0 = z_0$ for some constants $\theta, \beta > 0$ and $\alpha, z_0 \in \mathbb{R}$.

reasonable to expect that the setup we introduced in the opening of Chapter 5 qualitatively leads to the same results, although the structure of the noise is a little different. In order to rigorously justify all arguments, we prefer the setup we introduced earlier. We will see that the restriction of $G = \mathrm{SU}(d)$ is actually not needed, even though this is explicitly used in the discussion in [19], we comment on this in Remark. 7.10. As in Chapter 5, we denote the drift as $x = iH$, the controls by $y_j = iV_j$ and with $X = x^l, Y_j = y_j^l$ the respective left-invariant vector fields, which are simply given by $X(U_t) = x^l(U_t) = iU_tH$ in the standard representation. Throughout, we will be concerned with a diffusion process over a subgroup $G \subset \mathrm{SU}(d)$, modelled as the solution to the left-invariant Stratonovich differential equation

$$dU_t = X(U_t)dt + \sum_j Y_j(U_t) \odot dB_t^j, \quad U_0 = 1_G. \tag{7.2}$$

As a corollary to the convergence result Thm. 6.12 we immediately see that all moments of the induced distribution converge exponentially fast to the respective moments of the Haar measure for a weakly hypoelliptic diffusion.

Corollary 7.1. Let $(U_t)_t$ a weakly hypoelliptic diffusion process over $G \subset \mathrm{SU}(d)$, given as solution to (7.2) and let $\mu_t = Law(U_t)$ the associated convolution semigroup. Then for any (finite-dimensional) representation Φ of G, the Φ-moment of μ_t converges exponentially fast to the respective moment of the Haar measure as $t \to \infty$.

Proof. Since a representation $\Phi : G \to \mathrm{GL}(V)$ is Lipshitz with Lipshitz constant $\|\phi\|_{op}$, we can estimate the difference of the moments in norm for a suitable $C_\Phi > 0$ by

$$\|M_{\mu_t}^\Phi - M_G^\Phi\| = \left\| \int_G \Phi(U)d\mu_t - \int_G \Phi(U)d\mu_G \right\| \leq C_\Phi \mathcal{W}(\mu_t, \mu_G). \tag{7.3}$$

With the assumption on weak hypoellipticity, we can invoke Thm. 6.12 and conclude that there exists $\alpha > 0$ such that

$$\|M_{\mu_t}^\Phi - M_{\mu_G}^\Phi\| \leq C_\Phi e^{-t\alpha} \tag{7.4}$$

holds for all $t \geq 0$ and representations Φ, possibly after adapting $C_\Phi > 0$. Because we are working only in finite dimensions, (7.3) is valid in all norms up to a modification of the constant C_Φ. □

The task to explicitly determine the rate of convergence in this method is rather daunting, which is evident from the concluding discussion of Chapter 6. We therefore present an alternative approach to derive convergence of the moments, relying on properties of unitary representations. The obtained rates depend on the spectral gaps of ceratin operators. Computing these gaps is rather difficult, however these *can* in principle be computed, therefore we believe that the following discussion still serves a purpose.

7.2 Infinitesimal Generators

We know that the distribution $\mu_t := Law(g_t)$ of a Lévy process $(g_t)_t$ forms a continuous convolution semigroup if initially $\mu_0 = \delta_{1_g}$. It is easy to see that the moments taken with respect to any G-representation Φ then form a matrix semigroup $M^\Phi_{\mu_{t+s}} = M^\Phi_{\mu_t * \mu_s} = M^\Phi_{\mu_t} M^\Phi_{\mu_s}$. As we are only considering finite-dimensional representations, it is immediate that such a semigroup admits a generator $\mathcal{L}_\Phi$ in the sense that $M^\Phi_{\mu_t} = \exp(t\mathcal{L}_\Phi)$ for all $t \geq 0$. Recall that the probability operators P_t associated to the convolution semigroup μ_t are defined by $P_t f(g) = \mathbb{E}_{\mu_t}[l_g^* f]$ where l_g denotes the left-multiplication with $g \in G$ as usual. The moments $M^\Phi_{\mu_t}$ are obtained by applying the probability operator P_t to the coordinate functions of the representation: $M^\Phi_{\mu_t} = P_t\Phi$. It is reasonable to expect that we can canonically derive the generator $\mathcal{L}_\Phi$ from the infinitesimal generator $\mathcal{L}$ of the Lévy process. This is formally shown in the following proposition, thereby also proving equation (6) in [19] in a more general setting.

Proposition 7.2. Consider a Lévy process $(U_t)_t$ on a Lie group G with left-invariant infinitesimal generator $\mathcal{L} = X + \frac{1}{2}\sum_j Y_j^2$, or equivalently the solution to the left-invariant Stratonovich differential equation

$$dU_t = X(U_t)dt + \sum_j Y_j(U_t) \odot dB_t^j,\ U_0 = 1_G \tag{7.5}$$

for some independent Brownian motions $B^1, \ldots, B^m$. Let μ_t denote the law of U_t. Consider a finite-dimensional G-representation $\Phi : G \to Gl(V)$ and let $\phi : \mathfrak{g} \to \mathrm{End}(V)$ be the induced representation of the Lie algebra. Then the generator of the semigroup $\mathbb{E}_{\mu_t}[\Phi] = M^\Phi_{\mu_t}$ is given by

$$\mathcal{L}_\phi := \phi_x + \frac{1}{2}\sum_j \phi_{y_j}^2, \tag{7.6}$$

in the sense that $M^\Phi_{\mu_t} = \exp(t\mathcal{L}_\phi)$ holds for all times $t \geq 0$.

Proof. Since Φ is smooth, we have $\Phi \in \mathrm{Dom}(\mathcal{L})$ in the sense that this holds for all component functions. It is clear that $M^\Phi_{\mu_t} = \mathbb{E}_{\mu_t}[\Phi] = P_t\Phi(1_G)$. Using the Kolmogorov forward equation (5.9) for the transition semigroup P_t, we obtain the differential equation

$$\frac{d}{dt}M^\Phi_{\mu_t} = \frac{d}{dt}P_t\Phi = P_t\mathcal{L}\Phi(1_G).$$

We can compute $\mathcal{L}\Phi$ and use the left-invariance of $X, Y_1, \ldots, Y_m$ to arrive at

$$\mathcal{L}(\Phi) = X(\Phi) + \frac{1}{2}\sum_j Y_j^2(\Phi) = \left(\phi_x + \frac{1}{2}\sum_j \phi_{y_j}^2\right)\Phi.$$

Because $\phi_x + \frac{1}{2}\sum_j y_j^2$ acts linearly on the coordinates of Φ, it commutes with the expectation,

and we have

$$\frac{d}{dt}M^{\Phi}_{\mu_t} = \mathbb{E}_{\mu_t}[\mathcal{L}_\phi \Phi] = \mathcal{L}_\phi \mathbb{E}_{\mu_t}[\Phi] = \mathcal{L}_\phi M^{\Phi}_{\mu_t}. \tag{7.7}$$

Observing that initially $M^{\Phi}_{\mu_0} = M^{\Phi}_{\delta_{1_G}} = \mathbb{1}_V$, we clearly see that

$$M^{\Phi}_{\mu_t} = \exp(t\mathcal{L}_\phi), \tag{7.8}$$

which concludes the proof. □

Let us consider the special case where we have a control system $(iH|iV_1, \ldots, iV_m)$ on $\mathfrak{su}_d$ and the representation $\Phi = \Phi^{k,k}$. Here the generator $\mathcal{L}_k$ takes the form

$$\mathcal{L}_k(\rho) = i[H^{\otimes k}, \rho] - \frac{1}{2}\sum_j [V_j^{\otimes k}[V_j^{\otimes k}, \rho]] \text{ for } \rho \in \mathrm{End}(\mathbb{C}^d)^{\otimes k}, \tag{7.9}$$

where the tensor product $H \otimes H = \mathbb{1}_d \otimes H + H \otimes \mathbb{1}_d$ is to be understood in the Lie algebra sense. This generator is a special case of a *Lindblad operator*. In quantum dynamics, Lindblad operators generally refer to generators of quantum dynamical semigroups [75, 76], i.e. semigroups of quantum channels. In order to get an intuition as to why taking the expectation results in a quantum channel, we can approximate the expectation $\mathbb{E}[\rho_t] = \mathbb{E}_{\mu_t}[U\rho U^\dagger]$ for $t > 0$ by finite sampling of $U_1, \ldots, U_n \sim \mu_t$, so that

$$\mathbb{E}[\rho_t] \approx \frac{1}{n}\sum_{j=1}^{n} U_j \rho U_j^\dagger.$$

This corresponds to a measurement determined by the operators $\{1/\sqrt{n}U_1, \ldots, 1/\sqrt{n}U_n\}$ and therefore defines a quantum channel Λ_t, following Kraus Theorem 1.1. By taking the limit $n \to \infty$ we can heuristically argue that the operators Λ_t form a semigroup, and therefore equations along the lines of (7.8) can be expected. Treating the generators $\mathcal{L}_\phi$ within the framework of such quantum dynamical semigroups as in (7.9), adds no mathematically relevant structure for us, and rather distractes from the key arguments, therefore we leave it at this remark.

We continue with the observation that the generator $\mathcal{L}_\phi$ is just the action of the generator $\mathfrak{k} \in U(\mathfrak{g})$ in the universal envelopping algebra (see Rem. 5.7) under the representation ϕ. In order to clarify the relation between the different incarnations $\mathcal{L}$, we consider the commutative diagram:

$$\begin{array}{ccc} \mathfrak{k} \in U(\mathfrak{g}) & \xrightarrow{(\cdot)^l} & U(\mathrm{VF}^l(G)) \ni \mathcal{L} \\ \big\downarrow \phi & & \big\downarrow D\Phi \\ \mathcal{L}_\phi = \phi_{\mathfrak{k}} \in \mathrm{End}(V) & \xrightarrow{(\cdot)^l} & C^\infty(G_\Phi, \mathrm{End}(V)) \ni \mathcal{L}(\Phi) \end{array} \tag{7.10}$$

The horizontal maps are the representations of the Lie algebras by left-invariant derivations acting

on $C^\infty(G)$ and $C^\infty(\Phi(G))$ respectively. Since Φ is a group representation and therefore compatible with the group action, the vertical map on the right side is just the G-equivariant version of the Lie algebra representation on the left vertical.

The convergence of the Φ-moment semigroup is now encoded in the projected generator $\mathcal{L}_\phi$. In this context the spectral gap of $\mathcal{L}_\phi$ plays an important role. Recall that for self-adjoint operators, the spectral gap is the modulus of the smallest non-zero eigenvalue. Since for us $\mathcal{L}_\phi$ is in general *not* self-adjoint, we define the spectral gap $\Delta\mathcal{L}_\phi$ of $\mathcal{L}_\phi$ by

$$\Delta\mathcal{L}_\phi := \sup_{0\neq\lambda\in\Lambda_\phi} |\Re(\lambda)|, \tag{7.11}$$

where Λ_ϕ denotes the set of eigenvalues of $\mathcal{L}_\phi$. With this definition $\Delta\mathcal{L}_\phi$ can in principle be zero! We continue our discussion by collecting some basic algebraic properties of $\mathcal{L}_\phi$.

Lemma 7.3. Let $\mathfrak{g}$ be a compact Lie algebra and $\phi : \mathfrak{g} \to \mathrm{End}(\mathcal{H})$ a representation on a complex Hilbert space $\mathcal{H}$ which we assume to be unitary without loss of generality. Let $x, y_1, \ldots, y_m$ generate $\mathfrak{g}$ as a Lie algebra and consider the generator $\mathcal{L}_\phi = \phi_x + \frac{1}{2}\sum_j \phi_{y_j}^2$ as above. Then the following statements hold:

1. $\Re(\lambda) \leq 0$ for all eigenvalues λ of $\mathcal{L}_\phi$.
2. We have $\ker(\mathcal{L}_\phi) = \ker(\mathcal{L}_\phi^\dagger) = \mathcal{H}^\mathfrak{g} = \{v \in \mathcal{H} | \phi_z v =)0 \text{ for all } z \in \mathfrak{g}\}$.
3. Zero is a non-degenerate eigenvalue, i.e. if $\mathcal{L}_\phi^2 w = 0$ we already have $\mathcal{L}_\phi w = 0$.

Proof. First of all, because the representation is unitary - i.e. $\phi^\dagger = -\phi$ - we can write

$$\mathcal{L}_\phi = \phi_x + \frac{1}{2}\sum_j \phi_{y_j}^2 = \phi_x - \frac{1}{2}\sum_j \phi_{y_j}^\dagger \phi_{y_j}. \tag{7.12}$$

In particular, $\sum_j \phi_{y_j}^\dagger \phi_{y_j}$ a positive self adjoint operator.
Claim (1): For any eigenvalue λ and a corresponding normalised eigenvector w we have:

$$\lambda = \langle w, \mathcal{L}_\phi w\rangle = \underbrace{\langle w, \phi_x w\rangle}_{\in i\mathbb{R}} - \frac{1}{2}\sum_j \underbrace{\langle \phi_{y_j} w, \phi_{y_j} w\rangle}_{\in\mathbb{R}},$$

since ϕ_x is skew-adjoint. Therefore we get

$$\Re(\lambda) = -\frac{1}{2}\sum_j \langle \phi_j w, \phi_j w\rangle \leq 0.$$

Claim (2): For $w \in \ker(\mathcal{L}_\phi)$ we immediately see

$$0 = \Re(\langle w, \mathcal{L}_\phi w\rangle) = \frac{1}{2}\sum_j \langle \phi_{y_j} w, \phi_{y_j} w\rangle.$$

Therefore we have $w \in \ker(\phi_{y_j})$ for all j, and $\mathcal{L}_\phi w = \phi_x w = 0$ follows. Since $x, y_1, \ldots, y_m$ generate $\mathfrak{g}$ as a Lie algebra, we then have $\ker(\mathcal{L}_\phi) = \mathcal{H}^{\mathfrak{g}}$. For $\mathcal{L}^\dagger = -\phi_x + \frac{1}{2}\sum_j \phi_{y_j}^2$, we can use the same arguments after replacing x by $-x$.
Claim (3): Take $w \in \ker(\mathcal{L}_\phi^2)$. As we just saw, $\ker(\mathcal{L}_\phi) = \ker(\mathcal{L}_\phi^\dagger)$ and hence $\mathcal{L}_\phi^\dagger \mathcal{L}_\phi w = 0$. It is then straightforward to see

$$\|\mathcal{L}_\phi w\|^2 = \langle \mathcal{L}_\phi^\dagger \mathcal{L}_\phi w, w\rangle = 0,$$

and we have indeed $w \in \ker(\mathcal{L})$. □

Claims (2) and (3) lead to the following result.

Corollary 7.4. In the situation of the previous Lemma, let $w \in \ker(\mathcal{L}_\phi) \cap \mathrm{im}(\mathcal{L}_\phi)$. Then $w = \mathcal{L}_\phi v$ for some $v \in \mathcal{H}$ and $0 = \mathcal{L}_\phi(w) = \mathcal{L}_\phi^2(v)$. Following Claim (3) of Lemma 7.3, we get $w = \mathcal{L}_\phi v = 0$. Since $\mathcal{H}$ is finite dimensional, we deduce that there is a direct sum decomposition

$$\mathcal{H} = \ker(\mathcal{L}_\phi) \oplus \mathrm{im}(\mathcal{L}_\phi). \tag{7.13}$$

With Claim (2), and the well-known relations between kernels and images of adjoint linear maps, we see that $\ker(\mathcal{L}_\phi) = \ker(\mathcal{L}_\phi^\dagger) = \mathrm{im}(\mathcal{L}_\phi)^\perp$, and thus the decomposition (7.13) is even *orthogonal*.

Given some G-representation Φ, the first claim of Prop. 7.3 essentially confirms what we already know, namely that the moments are bounded and non-increasing. In order to show convergence of convolution semigroups $\mu_t = Law(g_t)$ with respect to the Φ-moment, we additionally need to ensure that $\mathcal{L}_\phi$ has no imaginary eigenvalues, as these will cause oscillating orbits of $M_{\mu_t}^\Phi = \exp(t\mathcal{L}_\phi)$, and therefore destroy convergence. This is accomplished by the following proposition, which implicitly assumes weak hypoellipticity of the process $(U_t)_t$, more precisely the equivalent Lie-theoretic characterisation in terms of regular control systems 6.5.

Proposition 7.5. Let $(x|y_1, \ldots, y_m)$ be regular generating set (see Def. 2.8) of a compact Lie algebra $\mathfrak{g}$ and $\phi : \mathfrak{g} \to \mathrm{End}(\mathcal{H})$ a unitary ($\phi = -\phi^\dagger$) Lie algebra representation on a Hilbert space $\mathcal{H}$. Then apart from $\lambda = 0$, the generator

$$\mathcal{L}_\phi = \phi_x + \frac{1}{2}\sum_j \phi_{y_j}^2$$

has no imaginary eigenvalues. In particular, the spectral gap of $\mathcal{L}_\phi$ is strictly positive $\Delta\mathcal{L}_\phi > 0$.

Proof. Recall that the regularity condition expresses that the zero-time ideal $\mathfrak{I}_0$ derived from the left-invariant control system $(x|y_1,\dots,y_m)$ satisfes $\mathfrak{I}_0 = \mathfrak{g}$. Assume that the generator $\mathcal{L}_\phi$ has a purely imaginary eigenvalue $\lambda \neq 0$ and let $w \in \mathcal{H}$ be a corresponding normalised eigenvector. Consider the Lie subalgebra of $\mathfrak{g}$ defined by

$$\mathrm{Ann}_\phi(w) = \{z \in \mathfrak{g} : \phi_z(w) = 0\},$$

which we call the *annihilator* of w in $\mathfrak{g}$ under the representation ϕ. Our goal is to show that $\mathrm{Ann}_\phi(w) = \mathfrak{g}$. We have $\lambda = \langle w, \phi_x w\rangle + \frac{1}{2}\sum_j \langle w, \phi_{y_j}^2 w\rangle$ where again the first summand is real and the second imaginary. Since λ is imaginary, and the ϕ_{y_j} are skew-adjoint we conclude

$$\sum_j \langle w, \phi_{y_j}^2 w\rangle = -\sum_j \|\phi_{y_j} w\|^2 = 0.$$

Consequently we have $\phi_{y_1} w = \cdots = \phi_{y_m} w = 0$ and therefore $\phi_x w = \lambda w$. Because ϕ is compatible with the Lie brackets $[\cdot,\cdot]$ we see that $\phi_{[x,y_j]} w = \phi_{[y_i,y_j]} w = 0$ holds for all i, j and iteratively for all nested commutators of $x, y_1, \dots, y_m$. Since $\mathfrak{g}$ is generated as a Lie algebra by $x, y_1, \dots, y_m$, we conclude that w is an eigenvector of ϕ_z for all $z \in \mathfrak{g}$. In particular, we get $[\mathfrak{g},\mathfrak{g}] \subset \mathrm{Ann}_\phi(w)$, which shows that $\mathrm{Ann}_\phi(w) \subset \mathfrak{g}$ is actually an ideal:

$$[\mathrm{Ann}_\phi(w), \mathfrak{g}] \subset [\mathfrak{g},\mathfrak{g}] \subset \mathrm{Ann}_\phi(w) \lhd \mathfrak{g}. \tag{7.14}$$

Because $\phi_{y_1} w = \cdots = \phi_{y_m} w = 0$, we have $y_1, \dots, y_m \in \mathrm{Ann}_\phi(w)$, and therefore

$$\mathfrak{I}_0 \subset \mathrm{Ann}_\phi(w) \lhd \mathfrak{g}. \tag{7.15}$$

With the assumption that the zero-time ideal satisfies $\mathfrak{I}_0 = \mathfrak{g}$ we can conclude indeed that $\mathrm{Ann}_\phi(w) = \mathfrak{g}$. In particular we have $\phi_x w = 0$ and therefore $\lambda = 0$. □

It is not difficult to provide a counterexample in case the generating set is non-regular:

Proposition 7.6. If a left-invariant control system $(x|y_1,\dots,y_m)$ controllable on $G \subset \mathrm{SU}(d)$ does not satisfy the regularity assumption $\mathfrak{I}_0 = \mathfrak{g}$, there exists a representation Φ of G, such that the induced action of $\mathfrak{k} = x + \frac{1}{2}\sum_j y_j^2 \in U(\mathfrak{g})$ under ϕ has a non-zero imaginary eigenvalue.

Proof. Because the Lie algebra $\mathfrak{g}$ is compact, it is reductive and therefore splits into a direct sum of ideals

$$\mathfrak{g} = \mathfrak{z}(\mathfrak{g}) \oplus [\mathfrak{g},\mathfrak{g}], \tag{7.16}$$

which are its center and the derived algebra respectively (Prop. A.30). Because we have $[\mathfrak{g},\mathfrak{g}] \subset \mathfrak{I}_0$, this decomposition is nontrivial, if $\mathfrak{I}_0 \subsetneq \mathfrak{g}$. The assumption on $(x|y_1,\dots,y_m)$ violating the condition $\mathfrak{I}_0 = \mathfrak{g}$ implies that at least one generator τ of the center $\mathfrak{z}(\mathfrak{g})$ is not contained in $\mathfrak{I}_0 \supset [\mathfrak{g},\mathfrak{g}]$ by (7.16). Since $\langle\tau\rangle_{Lie} = \mathrm{span}_{\mathbb{R}}\{\tau\} \subset \mathfrak{g}$ is an ideal which is complementary to $\mathfrak{I}_0$,

we have another decomposition of $\mathfrak{g}$ into ideals

$$\mathfrak{g} = \langle \tau \rangle_{Lie} \oplus \mathfrak{I}_0 \tag{7.17}$$

Let $G_1 := G_\tau \subset \mathcal{Z}(G)$ and $G_2 := G_{\mathfrak{I}}$ be the subgroups of G generated by τ and $\mathfrak{I}_0$ respectively. Because the generating Lie subalgebras are ideals in $\mathfrak{g}$, we can write G as the direct product $G = G_1 \times G_2$ following Prop. A.6. Consider the representation Φ of G given by the projection on the G_1-component $g = (g_1, g_2) \mapsto g_1 \in \mathrm{SU}(d)$. Clearly the induced Lie algebra representation ϕ is then just the projection

$$\phi : \mathfrak{g} = \langle \tau \rangle_{Lie} \oplus \mathfrak{I}_0 \to \langle \tau \rangle_{Lie}.$$

Note that we also have $\tau \notin \mathfrak{g}_0 \subset \mathfrak{I}_0$ and since $\tau \notin [\mathfrak{g}, \mathfrak{g}]$, τ must originate from the drift x which therefore satisfies $\phi_x \neq 0$. By construction we have $\phi_{y_j} = 0$ for all the other generators $y_1, \ldots, y_m$. Since skew-adjoint endomorphism on $\mathbb{C}^d$ are diagonalisable with purely imaginary eigenvalues, $\mathcal{L}_\phi = \phi_x$ then indeed has at least one imaginary eigenvalue which is non-zero. □

Remark 7.7. Together with Corollary 7.1, this counterexample shows that in order to get convergence of the process in total variation distance, weak hypoellipticity is also *necessary*, thereby proving the converse direction statement of Thm. 6.12.

Example 7.8. We go back to example 2.12, where we considered the Bose-Hubbard model on two qubits. Recall that the drift Hamiltonian was

$$H_0 = \sigma_x^1 \sigma_x^2 + \sigma_y^1 \sigma_y^2 + \sigma_z^1 + \sigma_z^2$$

in this case, and $H_c = \sigma_x^1 \sigma_x^2$ is the control term. Then the generated Lie algebra $\mathfrak{g}$ is contained in $\mathfrak{su}_{16}$ and we choose $\Phi = \Phi^{2,2}$ to be the tensor square of the conjugate standard representation of $\mathrm{SU}(16)$. A computation in Magma [28] shows that in this case the eigenvalues of the Lindbladian $\mathcal{L}_\phi = \phi_{iH_0} + \frac{1}{2}\phi_{iH_c}^2$ are indeed all non-imaginary $\{-2, -2 \pm 4i, -8 \pm 8i, 0\}$.

7.3 Convergence

Before we proceed to show convergence of the moment semigroups, we can easily derive a monotonicity result similar to Eq. 6.5 for the trace norm of $M_{\mu_t}^\Phi$. As ususal let $\mathcal{L} = X + \frac{1}{2}\sum_j Y_j^2$ denote the generator of the left-inavriant diffusion process in question. If $\phi_{y_j} \neq 0$ holds for at

least one j (compare with Prop. 7.6 and Prop. 7.5), we have

$$\begin{aligned}\frac{d}{dt}\|M^{\Phi}_{\mu_t}\|_2^2 &= \frac{d}{dt}\operatorname{Tr}\left(\exp(t\mathcal{L}_\phi)\exp(t\mathcal{L}_\phi^\dagger)\right) = \operatorname{Tr}\left((\mathcal{L}_\phi+\mathcal{L}_\phi^\dagger)\exp(t\mathcal{L}_\phi)\exp(t\mathcal{L}_\phi^\dagger)\right)\\ &= -\sum_j \operatorname{Tr}\left(\phi_{y_j}^\dagger\phi_{y_j}\exp(t\mathcal{L}_\phi)\exp(t\mathcal{L}_\phi^\dagger)\right)\\ &= -\sum_j \|\phi_{y_j}\exp(t\mathcal{L}_\phi)\|_2^2 < 0,\end{aligned} \tag{7.18}$$

since $\exp(t\mathcal{L}_\phi)$ is an isomorphism. With the properties of the generator $\mathcal{L}_\phi$ we previously discussed, we can proceed to give a refined version of the proof presented in section B of [19]. Specifically, we can replace the requirement of *controllability* on $\mathfrak{su}_d$ by *regularity* of the system on $\mathfrak{g} \subset \mathfrak{su}_d$, thereby becoming much more general. As already outlined earlier, this includes all cases in which the generated Lie algebra $\mathfrak{g}$ is *semisimple*. The proof also links the rate of convergence to the spectral gap $\Delta\mathcal{L}_\phi$ which can in principle be computed, however scaling of dimensions does become a problem here.

Theorem 7.9. Let $(g_t)_t$ be a weakly hypoelliptic diffusion process on a compact Lie group G with generator $\mathcal{L}$, described for instance by (7.2) with a regular left-invariant control system $(x|y_1,\dots,y_m)$. Then for any complex, finite dimensional G-representation Φ, the gap $\Delta\mathcal{L}_\phi$ of the projected generator generator is positive and there is a constant $C>0$ such that:

$$\|M^{\Phi}_{\mu_t} - M^{\Phi}_G\|_{op} \leq Ce^{-t\Delta\mathcal{L}_\phi/2} \text{ for all } t \geq 0. \tag{7.19}$$

In particular, μ_t is an ε-approximate Φ-Haar measure for all times

$$t \geq T(\varepsilon,\Phi,C) := \frac{2}{\Delta\mathcal{L}_\phi}\left(\log(1/\varepsilon)+\log(C)\right). \tag{7.20}$$

Proof. Fix a representation $\Phi : G \to \mathrm{GL}(\mathcal{H})$ onto a d-dimensional Hilbert space $\mathcal{H}$, which we assume to be unitary without loss of generality. We will proceed as follows: first we show that the invariant subspaces of $M^{\Phi}_{\mu_t} = \exp(t\mathcal{L}_\phi)$ and M^{Φ}_G coincides, and then prove that for any element w in their orthogonal complement the difference vanishes in the limit $t \to \infty$.
By Prop. 3.3 M^{Φ}_G is the orthogonal projection operator onto the space $\mathcal{H}^G$ invariant under the G-action. Because G is connected, this coincides with the invariants $\mathcal{H}^{\mathfrak{g}}$ of the Lie algebra representation, which happens to be the kernel of $\mathcal{L}_\phi$ by Prop 7.3. In particular $\mathcal{H}^G$ is invariant under $M^{\Phi}_{\mu_t} = \exp(t\mathcal{L}_\phi)$ as well. Let $v^{\|}$ and $v^{\perp}$ denote the parts of $v \in \mathcal{H}$ parallel and orthogonal to $\ker(\mathcal{L}_\phi) = \mathcal{H}^{\mathfrak{g}}$ respectively. We then have that

$$\left(M^{\Phi}_{\mu_t} - M^{\Phi}_G\right)v = \exp(t\mathcal{L}_\phi)v - M^{\Phi}_G v = \exp(t\mathcal{L}_\phi)v^{\perp}. \tag{7.21}$$

By Corollary 7.4 we can decompose $\mathcal{H}$ into an orthogonal sum via

$$\mathcal{H} = \ker(\mathcal{L}_\phi) \oplus \operatorname{im}(\mathcal{L}_\phi). \tag{7.22}$$

In particular $\mathcal{L}_\phi$ restricts to an isomorphism $\operatorname{im}(\mathcal{L}_\phi) \to \operatorname{im}(\mathcal{L}_\phi)$ which we will denote by $\widehat{\mathcal{L}}_\phi$ in the following. All that is left to do is to show that $\|\exp(t\widehat{\mathcal{L}}_\phi)\|_{op} \to 0$ as $t \to \infty$. At the core this relies on Prop. 7.5 which guarantees that we have $\mathcal{L}_\phi > 0$ if we assume that the process is weakly ypoelliptic. Let $\widehat{\mathcal{L}}_\phi = \widehat{\mathcal{L}}_\phi^S + \widehat{\mathcal{L}}_\phi^N$ denote the Jordan-Chevalley decomposition (see Prop. A.31) of $\widehat{\mathcal{L}}_\phi$ into its commuting semisimple (diagonalisable) and nilpotent parts respectively. Since $\widehat{\mathcal{L}}_\phi^S$ and $\widehat{\mathcal{L}}_\phi^N$ commute and $\widehat{\mathcal{L}}_\phi$ and $\widehat{\mathcal{L}}_\phi^S$ have the same eigenvalues, we get

$$\|\exp(t\widehat{\mathcal{L}}_\phi)\|_{op} \leq \|\exp(t\widehat{\mathcal{L}}_\phi^S)\|_{op}\|\exp(t\widehat{\mathcal{L}}_\phi^N)\|_{op} \leq e^{-t\Delta\mathcal{L}_\phi}\|\exp(t\widehat{\mathcal{L}}_\phi^N)\|_{op}. \tag{7.23}$$

Since $\widehat{\mathcal{L}}_\phi^N$ is nilpotent, we find $n \in \mathbb{N}$ such that $\left(\widehat{\mathcal{L}}_\phi^N\right)^{n+1} = 0$. Then we have

$$\|\exp(t\widehat{\mathcal{L}}_\phi^N)\|_{op} \leq \sum_{j=0}^{n} \frac{(t\|\widehat{\mathcal{L}}_\phi^N\|_{op})^j}{j!} \leq \|\widehat{\mathcal{L}}_\phi^N\|_{op}^n \max\{1, t^n\}. \tag{7.24}$$

In order to control the growth of the right hand side, we sacrifice a part of the exponential decay in (7.23). With some straightforward analysis we can bound

$$t^n e^{-\Delta\mathcal{L}_\phi t/2} \leq \left(\frac{2n}{\Delta\mathcal{L}_\phi}\right)^n e^{-n} \leq \left(\frac{n}{\Delta\mathcal{L}_\phi}\right)^n, \tag{7.25}$$

which we assume to be ≥ 1 without loss of generality. If we insert first (7.24) and then (7.25) into (7.23) we arrive at

$$\|\exp(t\widehat{\mathcal{L}}_\phi)\|_{op} \leq \left(\frac{n\|\widehat{\mathcal{L}}_\phi^N\|_{op}}{\Delta\mathcal{L}_\phi}\right)^n e^{-t\Delta\mathcal{L}_\phi/2}. \tag{7.26}$$

Becasue $\Delta\mathcal{L}_\phi > 0$, the right hand side vanishes in the limit $t \to \infty$ which finishes the proof. □

Remark 7.10. In order to compute the invariant space of the G-action, the authors of [19] assumed that everything is happening on some $\mathrm{SU}(d)$. In this case, the invariants of the representation $\Phi^{k,k}$ are just permutations of tensor factors as a consequence of the Schur-Weyl duality, cf. [77]. Subsequently they identified the projection onto this subspace with the limit of $\exp(t\mathcal{L}_\phi)$ drawing from other sources such as [78, 79]. Backtracking all the arguments it becomes clear that this argumentation contains unnecessary identifications, which allows us to lift the restriction that $G = \mathrm{SU}(d)$.

Remark 7.11. In the proof of the previous theorem we could also argue as follows: With weak hypoellipticity we can conclude that the process has a density, using Prop. 6.5. Hence the support of μ_t becomes dense for sufficiently large times, as implicated by Lemma 6.9. With Lemma 5.3 in

[17], the difference of the moments in operatornorm then satisfies

$$\|M^{\Phi}_{\mu_t} - M^{\Phi}_G\|_{op} < 1 \text{ for any } t > 0. \tag{7.27}$$

As before, we can conclude that

$$\{v \in \mathcal{H} | M^{\Phi}_{\mu_t} v = v\} \supset \ker(\mathcal{L}_\phi) = \{v \in \mathcal{H} | M^{\Phi}_G v = v\}.$$

With the decomposition $\mathcal{H} = \ker(\mathcal{L}_\phi) \oplus \ker(\mathcal{L}_\phi)^\perp$ it is easy to show that for any $m \in \mathbb{N}$:

$$\left(M^{\Phi}_{\mu_t} - M^{\Phi}_G\right)^m = \left(M^{\Phi}_{\mu_t}\right)^m - M^{\Phi}_G.$$

Since the composition of the moments is $\left(M^{\Phi}_{\mu_t}\right)^m = M^{\Phi}_{\mu_t^{*m}}$, we can write

$$\|M^{\Phi}_{\mu_{mt}} - M_G\|_{op} = \| \left(M^{\Phi}_{\mu_t} - M^{\Phi}_G\right)^m \|_{op} \leq \|M^{\Phi}_{\mu_t} - M^{\Phi}_G\|^m_{op}$$

which converges to zero as $m \to \infty$ because of (7.27). The problem here is again, that we cannot really determine the rate of decay more explicitly.

How does the spectral gap $\Delta\mathcal{L}_\phi$ behave in terms of the chosen representation, and more importantly what happens in a setup that can be scaled, e.g. for the Hubbard-model on a chain, which we discussed in the introduction? We basically already answered the first question with the introductory corollary.

Corollary 7.12. For a weakly hypoelliptic Lévy process with generator $\mathcal{L}$ we have

$$\inf_\phi \Delta\mathcal{L}_\phi > 0.$$

Proof. Let $\widehat{\mathcal{L}}_\phi$ denote the restriction of $\mathcal{L}_\phi$ onto $\ker(\mathcal{L}_\phi)^\perp$ as in the proof of Thm. 7.9. By Cor. 7.1, we know that there exists a constant $\alpha > 0$ such that

$$\| \exp(t\widehat{\mathcal{L}}_\phi)\|_{op} = \|M^{\Phi}_{\mu_t} - M^{\Phi}_G\|_{op} \leq C_\Phi e^{-\alpha t} \tag{7.28}$$

holds for all representations Φ and $t \geq 0$ with some suitably chosen $C_\Phi > 0$. Fix a representation Φ and let $\lambda_1 \ldots, \lambda_m$ denote the eigenvalues of $\widehat{\mathcal{L}}_\phi$. Then $e^{t\lambda_1}, \ldots, e^{t\lambda_m}$ are the eigenvalues of $\exp(t\widehat{\mathcal{L}}_\phi)$ and therefore $\| \exp(t\widehat{\mathcal{L}}_\phi)\|_{op} \geq e^{-t\Delta\mathcal{L}_\phi}$. In particular we have

$$e^{-t\Delta\mathcal{L}_\phi} \leq C_\Phi e^{-\alpha t}$$

for all $t \geq 0$ which implies that $\Delta\mathcal{L}_\phi \geq \alpha$ needs to hold for all G-representations Φ. □

Naturally, the more interesting problem is the harder one to deal with. Clearly the answer to the second question is depending on the structure of the chosen system, and we can give no general

answer. It would be interesting however to investigate if one can connect the spectral gaps of the operator $\mathcal{L}_\phi$ to the gaps of the control and drift Hamiltonians, even if only through crude estimates. Because the groundstate of the quantum system state should exhibit certain inertia in order for the quantum processor to work properly, the gaps of the drift and control Hamiltonians are required not to become too small in applications. Due to the non-commutativity of the generators, finding such estimates is expected to be rather difficult, if possible at all. Even if we cancontrol the exponential rate in some way, Thm. 7.9 is again of limited practical applicability, since we have no control over the operator norm of the nilpotent part $\widehat{\mathcal{L}}_\phi^N$ in general.

Interestingly, the convergence in distribution of a Lévy process on a compact semisimple Lie group $G \subset \mathrm{SU}(d)$ to the Haar measure μ_G is completely encoded in second moment.

Proposition 7.13. Assume that we have a control system $(x|y_1, \ldots, y_m)$ on a compact Lie group $G \subset \mathrm{SU}(d)$ (no assumptions on controllability!) such that the diffusion $(U_t)_t$ generated by $\mathcal{L} = X + \frac{1}{2}\sum_j Y_j^2$ converges to the Haar measure μ_G in the $\Phi^{2,2}$ moment. Then we also have convergence to the Haar measure in total variation, and in particular with respect to all other moments.

Proof. Let $V := \mathrm{End}(\mathbb{C}^d)^{\otimes 2}$ denote the complex Hilbert space on which $\Phi^{2,2}$ acts and let $\mathfrak{h} \subset \mathfrak{g}$ be the Lie subalgebra generated by $x, y_1, \ldots, y_m$. The moment semigroup $M^2_{\mu_t} = \exp(t\mathcal{L}_2)$ of $\mu_t = Law(U_t)$ is invariant on the subspace $\ker(\mathcal{L}_2) = V^{\mathfrak{h}} = \mathfrak{com}_{\phi\otimes\phi}(\mathfrak{h})$ where ϕ denotes the standard representation of $\mathrm{SU}(d)$. The moment M^2_G on the other hand is the orthogonal projection onto the subspace $V^{\mathfrak{g}} = \mathfrak{com}_{\phi\otimes\phi}(\mathfrak{g})$. In order to get convergence of $M^2_{\mu_t}$ to M^2_G, we must necessarily have

$$\mathfrak{com}_{\phi\otimes\phi}(\mathfrak{h}) = \mathfrak{com}_{\phi\otimes\phi}(\mathfrak{g}). \tag{7.29}$$

Because $\mathfrak{g}$ is semisimple and the standard representation ϕ is faithful, it follows from Thm. 2.13 that (7.29) is equivalent to having $\mathfrak{g} = \mathfrak{h}$. In particular, the system is already controllable on G and therefore generates a weakly hypoelliptic Lévy process (since G is semisimple) for which we proved convergence in total variation in Chapter 5.

□

On more general groups, this observation is no longer true. In view of Corollary 7.1 convergence in the $\Phi^{2,2}$-moment is certainly necessary for convergence in total variation. However one can find examples where the process is not weakly hypoelliptic but the generator of the $\Phi := \Phi^{2,2}$-moment semigroup still converges. This can happen if the center of the image $\mathfrak{z}(\phi^{\otimes 2}(\mathfrak{g})) \subset \mathrm{End}(\mathbb{C}^d)^{\otimes 2}$ and the center $\mathfrak{z}(\mathfrak{com}_{\phi\otimes\phi}(\mathfrak{g})) \subset \mathrm{End}(\mathbb{C}^d)^{\otimes 2}$ of its commutant have non-empty intersection, which allows to find generators where $\mathcal{L}_2$ has no imaginary eigenvalues. Such cases exist, finding them is somewhat uninstructive though, so we leave it at this remark.

APPENDIX

A. LIE GROUPS, LIE ALGEBRAS AND THEIR REPRESENTATIONS

In the following, we will provide a short summary of results and definitions related to Lie algebras and their representation that are relevant to us.

A.1 Lie Groups and Algebras

We assume basic familiarity with Lie groups G and Lie algebras $\mathfrak{g}$, how to derive Lie algebras from Lie groups and construction of the exponential map $\exp : \mathfrak{g} \to G$. These can be found in Chapter II.8 of [77].
For a Lie group G, here we denote the identity element by e and the induced Lie algebra by $\mathfrak{g} = T_eG$. In the main part of the thesis we use 1_G instead in order to avoid confusion with exponentials. The left and right multiplication with $g \in G$ are denoted by l_g and r_g respectively. Recall that elements in $\mathfrak{g}$ canonically induce left- and right-invariant vector fields X^l, X^r on G by setting $X^l(g) = Dl_g x$, and $X^r(g) = Dr_g x$ respectively. If $G \subset \mathrm{GL}(V)$ is a matrix Lie group, and $\mathfrak{g} \subset \mathrm{End}(V)$ the corresponding matrix Lie algebra, the invariant vector fields can be expressed by

$$X^l(g) = gx, \quad X^r(g) = xg$$

using the algebra structure of $\mathrm{End}(V)$. The first few result clarify the relation between homoorphisms of Lie groups and Lie algebras

Theorem A.1 (Thm 8.44 , Thm. 20.19 in [74])**.** Let G, H be Lie groups with Lie algebras $\mathfrak{g}, \mathfrak{h}$ respectively.

1. Let $\Phi : G \to H$ be a Lie group homomorphism. Then the differential $\phi = D_e\Phi : T_eG = \mathfrak{g} \to \mathfrak{h} = T_eH$ is a Lie algebra homomorphism, that is $\phi([x, y]) = [\phi_x, \phi_y]$ holds for all $x, y \in \mathfrak{g}$.

2. Assume that G is simply connected. Then for every Lie algebra homomorphism $\phi : \mathfrak{g} \to \mathfrak{h}$ there is a unique homomorphism of Lie groups $\Phi : G \to H$ with $\phi_x = D_e\Phi(x)$.

Proposition A.2 (Prop. 20.8 in [74])**.** Let $\Phi : G \to H$ a homomorphism of Lie groups and $\phi = D_e\Phi : \mathfrak{g} \to \mathfrak{h}$ its infinitesimal version. Then for $x \in \mathfrak{g}$ we have:

$$\Phi \circ \exp_{\mathfrak{g}}(x) = \exp_{\mathfrak{h}} \circ \phi(x).$$

The next two theorems together show that every real Lie algebra can be realised as the Lie algebra to some matrix Lie group G

Theorem A.3 (Ado's Theorem, Thm. E.4 in [77])**.** Every finite dimensional Lie algebra $\mathfrak{g}$ admits a faithful finite-dimensional representation, i.e. $\mathfrak{g} \subset \mathrm{End}(V)$ for some finite-dimensional vector space V.

Theorem A.4 (Lie's third fundamental Theorem, Thm. 20.22 in [74])**.** For every every finite-dimensional real Lie algebra there exists a uniquely determined simply connected Lie group the Lie algebra of which is $\mathfrak{g}$.

Definition A.5. Let $\mathfrak{g}$ be a Lie algebra. A Lie subalgebra $\mathfrak{h} \subset \mathfrak{g}$ satisfying $[\mathfrak{h}, \mathfrak{g}] \subset \mathfrak{h}$ is called an *ideal* in $\mathfrak{g}$, usually written $\mathfrak{h} \lhd \mathfrak{g}$. This is the condition sufficient and necessary to obtain a well-defined bracket on quotient $\mathfrak{g}/\mathfrak{h}$ which is then again a Lie algebra.

Proposition A.6 (Follows from Thm. 20.28 in [74])**.** Let G be a Lie group and $G_1, G_2 \subset G$ Lie-subgroups. Denote their respective Lie algebras by $\mathfrak{g}, \mathfrak{g}_1$ and $\mathfrak{g}_2$ respectively. Then G can be decomposed as a direct product $G = G_1 \times G_2$ if and only if $\mathfrak{g}_1, \mathfrak{g}_2 \lhd \mathfrak{g}$ are ideals and satisfy $\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2$.

Remark A.7. Every Lie algebra $\mathfrak{g}$ has canonically two special ideals, the *derived Lie algebra* $[\mathfrak{g}, \mathfrak{g}]$, and the *center* $\mathfrak{z}(\mathfrak{g}) = \{x \in \mathfrak{g} | [x, y] = 0 \text{ for all } y \in \mathfrak{g}\}$.

Definition A.8. A non-abelian Lie algebra $\mathfrak{g}$ is called *simple* if it has only the tautological ideals $\{0\}, \mathfrak{g} \lhd \mathfrak{g}$. It is called *semisimple* if it admits a decomposition into pairwise commuting simple subalgebras.

In particular, every semisimple Lie algebra satisfies $\mathfrak{g} = [\mathfrak{g}, \mathfrak{g}]$ and $\mathfrak{z}(\mathfrak{g}) = \{0\}$. A Lie algebra satisfying $[\mathfrak{g}, \mathfrak{g}] = \mathfrak{g}$ is called *perfect*. The converse conclusion that a perfect Lie algebra is semisimple is in general not satisfied.

Every Lie algebra carries a canonical bilinear form, which encodes several interesting properties of the Lie algebra.

Definition A.9. We define the *Killing form* of $\mathfrak{g}$ to be the bilinear form

$$\mathcal{K}(x, y) := \mathrm{Tr}(\mathrm{ad}_x \, \mathrm{ad}_y),$$

where $\mathrm{ad}_x \in \mathrm{End}(\mathfrak{g})$ denotes the map $y \mapsto [x, y]$.

Proposition A.10 (Prop. C.10 in [77])**.** A Lie algbera $\mathfrak{g}$ is semisimple if and only if the Killing form $\mathcal{K}_\mathfrak{g}$ is nondegenerate.

Definition A.11 (Universal Envelopping algebra)**.** Let $\mathfrak{g}$ be a Lie algebra over $K = \mathbb{R}, \mathbb{C}$. The universal envelopping algebra is the algebra consisting of formal products $x_1 \dots x_n$ obeying the commutator relation $xy - yx = [x, y]$. Formally,

$$U(\mathfrak{g}) := T(\mathfrak{g})/\{x \otimes y - y \otimes x - [x, y]\}$$

where $T(\mathfrak{g}) = K \oplus \mathfrak{g} \oplus \mathfrak{g}^{\otimes 2} \oplus \dots$ is the tensor algebra of $\mathfrak{g}$.

Definition A.12 (Casimir element)**.** Let $\mathfrak{g}$ be a Lie algebra with an ad-invariant, non-degenerate bilinear form $\langle \cdot, \cdot, \rangle$. Choose a basis $x_1, \dots, x_n$ of $\mathfrak{g}$ and let $x^1, \dots, x^n$ the dual basis with respect to $\langle \cdot, \cdot \rangle$. Then the *Casimir element* of $\langle \cdot, \cdot \rangle$ is the the element of the universal envelopping algebra given by

$$C_\mathfrak{g} = \sum x_i x^i \in U(\mathfrak{g}).$$

In case $\mathfrak{g}$ is semisimple we always consider the Casimir element to be constructed with respect to the Killing form. Any Casimir element constructed in this fashion is contained in the center of the universal envelopping algebra $U(\mathfrak{g})$.

A.2 Representation Theory

In everything that follows, vector spaces considered are considered to be complex and finite-dimensional unless indicated differently. As a general reference for this section we refer to [77, 74].

Definition A.13. Let V be a finite dimensional complex vector space and G a Lie group. A Lie group representation of G on V is a smooth homomorphism of Lie groups $G \to \mathrm{GL}(V)$. Similarly, a Lie algebra representation of a Lie algebra $\mathfrak{g}$ on V is a homomorphism of Lie algebras $\mathfrak{g} \to \mathrm{End}(V)$. An injective representation is called *faithful*.

By Φ_g, ϕ_x and Φ^v, ϕ^v we denote the restriction of the maps $\Phi : G \times V \to V$, $\phi : \mathfrak{g} \times V \to V$ to the second and first argument, respectively.
As a special case of Thm. A.1, representations of Lie groups and algebras are closely related:

Corollary A.14 (To Theorem A.1)**.** Let G be a Lie group with Lie algebra $\mathfrak{g}$. Then

1. every Lie group representation $\Phi : G \to \mathrm{GL}(V)$ on a finite-dimensional vector space differentiates to a Lie algebra representation $\phi : \mathfrak{g} \to \mathrm{End}(V)$,
2. if G is simply connected, every Lie algebra representation of $\phi : \mathfrak{g} \to \mathrm{End}(V)$ uniquely integrates to a Lie group representation of G.

In the following, if we have representations $\Phi : G \to \mathrm{GL}(V)$ and $\phi : \mathfrak{g} \to \mathrm{End}(V)$ on a Lie group with Lie algebra $\mathfrak{g}$, we shall always understand them to be related in this particular fashion. Because we will mostly be concerned with the standard representation of compact Lie groups, this result allows us to focus entirely on the representation theory of Lie algebras which is very well understood in the case where $\mathfrak{g}$ is semisimple. Lie groups and their algebras admit natural representations on the Lie algebra.

Example A.15. The *Adjoint representation* of G is the representation $Ad : G \to \mathrm{GL}(\mathfrak{g})$ defined by

$$Ad_g x = \frac{d}{dt}\Big|_{t=0} g \exp(tx) g^{-1} = D_e \operatorname{conj}_g x,$$

where $conj_g : G \to G\, h \mapsto ghg^{-1}$ denote the conjugation with $g \in G$. The induced representation $D_e Ad : \mathfrak{g} \to \mathrm{End}(\mathfrak{g})$ is called the *adjoint representation* and is given by

$$\mathrm{ad}_x\, z = [x, z].$$

In the special case, where we have $G \subset \mathrm{GL}(V)$ for some finite dimensional vector space V, we get

$$Ad_\theta : \mathrm{End}(V) \to \mathrm{End}(V)\ x \mapsto \frac{d}{dt}\Big|_{t=0} \theta \circ \exp(tx) \circ \theta^{-1} = \theta \circ x \circ \theta^{-1} = \mathrm{conj}_\theta\, x.$$

Similarly, the adjoint representation of $\mathfrak{g} \subset \mathrm{End}(V)$ is the ordinary commutator of linear maps.

Example A.16. Consider a Lie group G and a Lie algebra $\mathfrak{g}$ with representations $\Phi : G \to Gl\,(V)$, $\phi : \mathfrak{g} \to \mathrm{End}\,(V)$. We can naturally construct representations on the r-fold tensor product $V^{\otimes r}$ by

$$\Phi^{\otimes r} : g \mapsto \Phi_g \otimes \ldots \otimes \Phi_g \tag{A.1}$$

$$\phi^{\otimes r} : x \mapsto \sum_{j=1}^{r} \mathbb{1}_V^{\otimes j-1} \otimes \phi_x \otimes \mathbb{1}_V^{\otimes r-j}. \tag{A.2}$$

If Φ differentiates to ϕ, then $\Phi^{\otimes r}$ differentiates to $\phi^{\otimes r}$, heuristically one should think of the Leibniz rule for differentiation here.

For Lie group and algebra representations, we can canonically identify the following subspaces.

Definition A.17. Let $\Phi : G \to \mathrm{GL}(V)$ a Lie group representation, and $\phi : \mathfrak{g} \to \mathrm{End}(V)$ be representations of a Lie group G and Lie algebra $\mathfrak{g}$ respectively. We define the invariant subspaces $V^G, V^{\mathfrak{g}} \subset V$ to be

$$V^G := \{v \in V : \Phi_g v = v \text{ for all } g \in G\},$$
$$V^{\mathfrak{g}} := \{v \in V : \phi_x v = 0 \text{ for all } x \in \mathfrak{g}\}.$$

Definition A.18. A Lie algebra representation $\phi : \mathfrak{g} \to \mathrm{End}(V)$ is called *irreducible* if it has precisely two subspaces $W_1, W_2 \subset V$ which are invariant under the action of $\mathfrak{g}$, that is ϕ restricts to a representation $\mathfrak{g} \to \mathrm{End}(W_i),\ i = 1, 2$. These are then of course $\{0\}$ and V. In particular, we do not consider $\{0\}$ to be irreducible.

Definition A.19. A representation of a Lie algebra is called *completely reducible* if it admits a direct sum decomposition into irreducible subrepresentations.

Theorem A.20 (Weyl, Thm. 9.19 in [77])**.** Every finite dimensional representation of a semisimple Lie algebra $\mathfrak{g}$ is completely reducible.

Due to the close relationship between Lie groups and their Lie algebras, the following is not surprising:

Proposition A.21. Let $\Phi : G \to \mathrm{GL}(V)$ be a representation of a Lie group and $\phi : \mathfrak{g} \to \mathrm{End}(V)$ the induced Lie algebra representation. If G is connected, then $V^G = V^{\mathfrak{g}}$.

Proof. Assume that $\Phi_g v = v$ for all g: Then $\phi_x v = D_e \Phi^v x = 0$. Conversely, if $\phi_x v = 0$ for all x, then

$$\Phi(\exp(x))v = \exp(\phi_x)v = \sum_n \frac{\phi_x^n}{n!} v = v,$$

according to the previous theorem. Because the image of $\mathfrak{g}$ under the exponential map $\exp(\mathfrak{g}) \subset G$ generates G as a Lie group, the claim follows. □

Definition A.22. Let $\phi : \mathfrak{g} \to V$ a Lie algebra representation on some vector space V. ϕ naturally induces a Lie algebra representation on $\mathrm{End}(V)$ by $x.\alpha = [\phi_x, \alpha]$ for $\alpha \in \mathrm{End}(V)$. We then define the *commutant* of $\mathfrak{g}$ to be the invariant subspace

$$\mathfrak{com}_{\mathfrak{g}}(V) := \mathrm{End}(V)^{\mathfrak{g}} = \{\alpha \in \mathrm{End}(V) | [\phi_x, \alpha] = 0 \text{ for all } x \in \mathfrak{g}\}.$$

For a Lie group representation $\Phi : G \to \mathrm{GL}(V)$, we similarly define the commutant by

$$\mathfrak{com}_G(V) := \mathrm{End}(V)^G = \{\alpha \in \mathrm{End}(V) | \Phi_g \alpha \Phi_g^{-1} = \alpha \text{ for all } g \in G\}.$$

These spaces are also referred to as the *symmetries* of the given representation. Of course, if the Lie algebra belongs to a connected Lie group, the two spaces coincide by the previous Proposition.

Remark A.23. One special class of representations that will occur frequently, are the representations $\Phi^{k,k}$ of $G \subset SU(d)$ on $\mathrm{End}(\mathbb{C}^d)^{\otimes k}$ defined by $\Phi_U^{k,k} : \rho \mapsto U^{\otimes k} \rho U^{\dagger \otimes k}$ for $k \geq 1$. We usually employ the Liouville space notation to write $\Phi_U^{k,k} = U^{\otimes k} \otimes \bar{U}^{\otimes k}$. Since $\Phi^{k,k}$ is just the

restriction of the k-th tensor power of the Adjoint representation of $\mathrm{GL}(\mathbb{C}^d)$, the induced Lie algebra representation $\phi : \mathfrak{g} \subset \mathfrak{su}_d \to \mathrm{End}(\mathrm{End}(\mathbb{C}^d)^{\otimes k})$ is given by $\phi_x(\rho) = [x^{\otimes k}, \rho]$ where the tensor product in the Lie algebra should be understood in the sense $x \otimes x = x \otimes \mathbb{1}_d + \mathbb{1}_d \otimes x$. Since we only consider subgroups of $\mathrm{SL}(\mathbb{C}^d)$, the representations $\Phi^{k,k}$ and $\phi^{k,k}$ are faithful for all $k \geq 1$.

A.3 The Compact Cases

The representation theory of Lie algebras which are not semisimple but arise from a compact Lie group can be essentially reduced to the semisimple case, as we will see in the following. The starting point for this is the following theorem.

Theorem A.24 (Haar measure - various results from Chp IV, §15 in [80])**.** On a compact Lie group G there is a unique probability measure μ_G, called the *Haar-measure* which is both left and right invariant, i.e. satisfies $\mu_G(gA) = \mu_G(Ag) = \mu_G(A)$ for all Borel sets $A \in \mathscr{B}(G)$, and $g \in G$. Additionally it is also *unimodular*, that is we have $\mu_G(A) = \mu_G(A^{-1})$ for all $A \in \mathscr{B}(G)$.

Proposition A.25. Every finite-dimensional complex representation $\Phi : G \to \mathrm{GL}(V)$ of a compact Lie group is unitary for a suitably chosen scalar product $\langle \cdot, \cdot, \rangle_\Phi$. In particular, the induced representation $\phi : \mathfrak{g} \to \mathrm{End}(V)$ is skew-hermitian, i.e. $\phi^\dagger = -\phi$ with resepct to $\langle \cdot, \cdot \rangle_\Phi$.

Proof. Pick any scalar product $\langle \cdot, \cdot \rangle$ on V and average over the Haar measue:

$$\langle x, y \rangle_\Phi := \int_G \langle \Phi_g x, \Phi_g y \rangle d\mu_g.$$

□

This shows that every compact Lie group can be realised as a subgroup of some unitary group $U(d)$. In particular, using the Adjoint/adjoint representations, the Lie algebra of a compact Lie group always carries an inner product, such that $\langle \mathrm{Ad}_g\, x, \mathrm{Ad}_g\, y \rangle = \langle x, y \rangle$ and $\langle [x, y], z \rangle = \langle x, [y, z] \rangle$ hold for all $x, y, z \in \mathfrak{g}$ and $g \in G$. Since the Adjoint representation can be decomposed as $\mathrm{Ad}_g = Dr_g^{-1} \circ Dl_g$ obtain the following corollary.

Corollary A.26. On a compact Lie group, there exists a bi-invariant Riemannian metric q_G such that the induced Riemannian volume measure coincides with the Haar measure.

Remark A.27. As one should expect from such a canonical construction, the metric q_G has a few properties which follow naturally, but nevertheless are important to observe. First of all, the unique geodesic starting in $g \in G$ with velocity $v \in T_g G$ is given by

$$\gamma_t = \exp(tDl_{g^{-1}} v)g. \tag{A.3}$$

Moreover, the induced Riemannian distance function $d(\cdot, \cdot)$ is bi-invariant, i.e.

$$d(gh_1, gh_2) = d(h_1, h_2) = d(h_1 g, h_2 g)$$

for all $g_1, \mathfrak{g}_s, h \in G$. If the minimising geodesic $\gamma : [0, 1] \to G$ between g_1 and g_2 is given by $\gamma_t = \exp(tx)g_1$, we have $d(g_1, g_2) = \|x\|_\mathfrak{g}$.

A necessary condition that a Lie algebra arises from a compact Lie group is captured by the next criterion and the following remark.

Definition A.28. A Lie algebra is called *compact* if its Killing form $\mathcal{K}$ is negative semidefinite.

Remark A.29. A compact Lie group always has a compact Lie algebra. Without loss of generality we can assume that the adjoint representation is skew-hermitian and therefore diagonalisable on the complexification $\mathfrak{g}_{\mathbb{C}} = \mathfrak{g} \otimes \mathbb{C}$ with imaginary eigenvalues. Thus ad_x^2 is a nonpositive operator for all $x \in \mathfrak{g}$ and therefore $\mathcal{K}(x,x) = \mathrm{Tr}(\mathrm{ad}_x^2) \leq 0$ for all $x \in \mathfrak{g}$. The converse statement is in general false! For example the Heisenberg group, which consists of upper triangular real 3×3 matrices with ones on the diagonal, is not compact but the corresponding Heisenberg algebra defined by the commutation relations $[x,y] = z$, $[x,z] = [y,z] = 0$ has negative semidefinite Killing form. However, we can find a compact Lie group G the Lie algebra of which is $\mathfrak{g}$ for every compact real Lie algebra $\mathfrak{g}$.

If we additionally assume a compact Lie algebra to be semisimple, the negative Killing form canonically provides us with an ad-invariant inner product on $\mathfrak{g}$. Due to the following result the representation theory of compact Lie algebras can completely be understood in terms of the semisimple case.

Proposition A.30. A compact real Lie algebra is *reductive*, that is its adjoint representation is completely reducible. In particular we have a decmposition into the central and derived ideals

$$\mathfrak{g} = \mathfrak{z}(\mathfrak{g}) \oplus [\mathfrak{g}, \mathfrak{g}], \tag{A.4}$$

and $[\mathfrak{g}, \mathfrak{g}]$ is semisimple. In particular a compact Lie algebra is semisimple if and only if $[\mathfrak{g}, \mathfrak{g}] = \mathfrak{g}$.

Proof. Since $\mathfrak{g}$ is the Lie algebra to some compact Lie group, it carries an invariant inner product $\langle \cdot, \cdot \rangle$. Let $\mathfrak{a} \subset \mathfrak{g}$ an invariant subspace, i.e. an ideal $[\mathfrak{g}, \mathfrak{a}] \subset \mathfrak{a}$. Then the orthogonal complement with respect to $\langle ., . \rangle_{\mathrm{ad}}$ is invariant as well, since $\langle [x, a'], a \rangle = -\langle a', [x, a] \rangle = 0$ for every $a' \in \mathfrak{a}^{\perp}$ and $a \in \mathfrak{a}$. Therefore the adjoint representation is completely reducible. Consider the decomposition

$$\mathfrak{g} = \mathfrak{c}_1 \oplus \cdots \oplus \mathfrak{c}_k \oplus \mathfrak{s}_1 \oplus \cdots \oplus \mathfrak{s}_m$$

of $\mathfrak{g}$ into commuting, irreducible subalgebras where $\mathfrak{c}_i$ are the one-dimensional summands spanning the center $\mathfrak{z}(\mathfrak{g})$. Clearly the irreducibility of the $\mathfrak{s}_j$ implies that these summands are *simple*. Hence the derived Lie algebra

$$[\mathfrak{g}, \mathfrak{g}] = \bigoplus \underbrace{[\mathfrak{c}_i, \mathfrak{c}_i]}_{=\{0\}} \oplus \bigoplus \underbrace{[\mathfrak{s}_j, \mathfrak{s}_j]}_{=\mathfrak{s}_j}.$$

is semisimple, and we arrive at (A.4). □

For the lack of a better opportunity, we state the following result here, which is the coordinate free version of the Theorem about Jordan normal forms.

Proposition A.31. Let $A \in \mathrm{End}(\mathbb{C}^d)$. Then there exist uniquely determined $S, N \in \mathrm{End}(\mathbb{C}^d)$ with $A = S + N$ such that S is diagonalisable, N nilpotent, and $[S, N] = 0$. Additionally A and S have the same eigenvalues up to multiplicities. This decomposition is usually called the *Jordan-Chevalley* decomposition.

B. ANALYSIS, PROBABILITY THEORY AND STOCHASTIC ANALYSIS

B.1 Functional Analysis

Definition B.1 (Distributions)**.** Consider an open set $\Omega \subset \mathbb{R}^n$ and let $\mathcal{D} := C_c^\infty(\Omega)$ be the space of compactly supported smooth functions. A *distribution* on Ω is an element of the dual space $\mathcal{D}'$ taken with respect to the L^2-inner product. This definition also transfers to Riemannian manifolds without any issues. The Schwartz space $\mathscr{S}(\Omega)$ is the space of rapidly decaying functions, that is $f \in \mathscr{S}(\Omega)$ if for each pair of multiindices α, β, we have

$$\lim_{x\to\infty} x^\alpha \partial^\beta f(x) = 0.$$

It is equipped with the topology induced by the family of seminorms

$$\|\varphi\|_N := \sup_{|\alpha|,|\beta|\leq N} \sup_x |x^\alpha \partial^\beta \varphi(x)|.$$

The topological dual $\mathscr{S}'(\Omega)$ is called the space of *tempered distributions*.

In general we will denote all pairings with dual objects and scalar products by $\langle \cdot, \cdot \rangle$.

Remark B.2. Since the inclusion $C_c^\infty(\Omega) \hookrightarrow C_c(\Omega)$ is continuous with dense image, the dual operator $C_c(\Omega)' \hookrightarrow C_c^\infty(\Omega)'$ is also a continuous injection. $C_c(\Omega)'$ can be identified with the space of Radon measures on Ω, which therefore are distributions in the functional analytic sense. In particular, probability measures are such Radon measures, thereby also justifying the commonly used term 'probability distribution'.

Lemma B.3. Let X be a Banach space and $f : \mathbb{R} \to X$ a map. Assume that f is weakly C^k, that is the map $t \mapsto \langle \alpha, f(t)\rangle$ is k times continuously differentiable for all dual elements $\alpha \in X'$. Then f is $k-1$ times strongly continuously differentiable.

Proof. At the core, the statement relies on the uniform boundedness principle, which states that a subset $S \subset X$ is weakly bounded, i.e.

$$\sup_{x\in S} \langle \alpha, x\rangle < \infty \text{ for all } \alpha \in X',$$

if and only if it is strongly bounded, i.e.

$$\sup_{x\in S} \|x\| \leq C \text{ for some } C < \infty,$$

cf. [62], Theorem 3.18. First we consider the cases where $k = 1, 2$, all others follow inductively. We start by considering a function f that is weakly C^1. We only show continuity around zero, all other points follow analogously. Because the limit

$$\lim_{s\to 0} s^{-1}\langle \alpha, f(s) - f(0)\rangle$$

exists for all $\alpha \in X'$, we can conclude that the map $\rho_\alpha(s,t) := \langle \alpha, (s-t)^{-1}(f(t)-f(s))\rangle$ is continuous for every $\alpha \in X'$ and therefore $\rho_\alpha([-\varepsilon,\varepsilon]^2)$ is bounded. Applying the uniform boundedness principle, we find $C>0$ such that $\|(s-t)^{-1}(f(s)-f(t))\| \leq C$ for all $s,t \in [-\varepsilon,\varepsilon]$. In particular, this shows that f is strongly continuous. The next step is to conclude that if f is weakly C^2 then it is strongly C^1. Because the first weak derivative f' is weakly C^1, it is continuous by the previous case. In particular, we can write $f(t) = f(t_0) + \int_{t_0}^t f'(s)ds$ for all $t \geq t_0 \in \mathbb{R}$, therefore f is differentiable with continuous derivative f' by the fundamental theorem of calculus. □

In order to show the same result for Frechét spaces - i.e. the topology originates from a family of *semi*norms - the argument does not fundamentally change, we just proceed by checking the claim seperately in the respective seminorms.

B.2 Basic Probability Theory

This part provides an overview of some results and definitions in probability theory and stochastic analysis, where we assume familiarity with the most fundamental related notions, otherwise we point at [81].

Miscellaneous

We will always denote probability measures either by the greek letters μ, ν, η or $\mathcal{P}$. General measurable spaces are denoted by $(S, \mathscr{A})$ and probability spaces by $(\Omega, \mathcal{P}, \mathscr{F})$. $\mathcal{P}$ denotes the probability measure, Ω the space of outcomes and $\mathscr{F}$ the $\sigma-$algebra of events. The set of probability measures on a measurable space S is denoted by $\mathcal{M}^1(S)$. If S is a topological space, we always consider the $\sigma-$algebra to be the respective Borel σ-algebra $\mathscr{A} = \mathscr{B}(S)$ of S.

For a random variable $X : \Omega \to S$ let $\mathcal{P}^X = Law(X) = X_*\mathcal{P}$ denote the induced probability measure on S given by $\mathcal{P}^X(A) = \mathcal{P}(X^{-1}(A))$ for all $A \in \mathscr{A}$. We write $X \sim \mu$ if $\mathcal{P}^X = \mu$. For a random variable $X : \Omega \to S$ with $X \sim \mu$ and a function $f : S \to \mathbb{R}^d$ we usually write $\mathbb{E}_\mu[f] = \int f d\mu = \mathbb{E}[f(X)] = \int f(X(\omega))d\mathcal{P}(\omega)$ for its expectation under μ. An important tool to estimate the expectation of random variables which iscentral to information theory is Jensen's inequality.

Proposition B.4 (Jensen's inequality)**.** For every random variable $X : \Omega \to \mathbb{R}$ and *convex* measurable function $f : \mathbb{R} \to \mathbb{R}$, we have $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$. If the function is concave, we get the reverse estimate.

In many situations it is necessary to determine the expectation of a random variable with respect to certain events which have measure zero, that are stochastically not independent.

Proposition B.5 (Def. 8.11 and Thm. 8.12 in [81])**.** Consider a probability space $(\Omega, \mathcal{P}, \mathscr{F})$ and a σ-subalgebra $\mathscr{G} \subset \mathscr{F}$. Let $X \in L^1(\mathcal{P})$ be an integrable random variable. Then there exists a unique $\mathscr{G}$-measurable random variable Y, such that

$$\mathbb{E}[\chi_S X] = \mathbb{E}[\chi_S Y]$$

for all $S \in \mathscr{G}$. Y is called the called the *conditional expectation of X with respect to $\mathscr{G}$* and denoted by $\mathbb{E}[X|\mathscr{G}] := Y$.

Proposition B.6 (Thms. 8.14 and 8.20 in [81])**.** Let $(\Omega, \mathcal{P}, \mathscr{F})$ be a probability space and $\mathscr{G} \subset \mathscr{F}$ a σ-subalgebra and X, Y real-valued random variables as above. Then the conditional expectation satisfies the following properties:

1. If X is $\mathscr{G}$-measurable, then $\mathbb{E}[X|\mathscr{G}] = X$
2. We have $\mathbb{E}[XY|\mathscr{G}] = \mathbb{E}[X]\mathbb{E}[Y|\mathscr{G}]$ if X independent of the sigma algebra $\sigma(Y, \mathscr{G}) \subset \mathscr{F}$ generated by Y and $\mathscr{G}$.
3. Law of total expectation: $\mathbb{E}[\mathbb{E}[X|\mathscr{G}]] = \mathbb{E}[X]$.
4. Jensen's inequality: $f(\mathbb{E}[X|\mathscr{G}]) \leq \mathbb{E}[f(X)|\mathscr{G}]$ for any convex function $f : \mathbb{R} \to \mathbb{R}$.

One possibility to characterise probability measures in vector spaces is to determine the integrals of polynomials.

Definition B.7. Let μ be a measure on an n-dimensional vector space V and $k \in \mathbb{N}$. Consider the inclusion map $\iota^k : V \to V^{\otimes k}$ defined by $v \mapsto v^{\otimes k}$. If for all multiindices $I \in \{1, \dots, n\}^k$ the integral $\int_V |x_{i_1}| \dots |x_{i_k}| d\mu(x)$ is finite, we say that the measure is of order k and define the k-th *moment* of μ by

$$M_\mu^k := \int_V \iota^k(x) d\mu(x) \in V^{\otimes k}.$$

The following proposition shows that a probability measure on a compact set is completely determined by its moments.

Proposition B.8. A probability measure on a compact subset $K \subset \mathbb{R}^d$ is completely determined by its moments, i.e. for two measures μ, ν we have $\mu = \nu$ if and only if $M_\mu^k = M_\nu^k$ holds for all $k \geq 0$.

Proof. Every monomial of order k in the coordinates $x_1, \dots, x_n$ corresponds to a component of the tensor power $x^{\otimes k}$. If we have $M_\mu^k = M_\nu^k$ for all orders $k \geq 0$, we can conclude that $\int Q(x_1, \dots, x_n) d\mu(x) = \int Q(x_1, \dots, x_n) d\nu(x)$ holds for all multivariate polynomials Q. Because K is compact, these polynomials are dense in the continuous functions $C(K)$ by the Stone-Weierstraß Theorem and therefore we have $\int f(x) d\mu(x) = \int f(x) d\nu(x)$ for all $f \in C(K)$. This is sufficient to conclude that $\mu = \nu$, following Theorem 2.14 in [82]. □

Definition B.9. For two probability measures μ, ν on a (semi-)group G, we define the convolution $\mu * \nu$ by

$$(\mu * \nu)(A) := \int \int \chi_A(gh) d\mu(g) d\nu(h)$$

for every Borel set $A \in \mathcal{B}(G)$.

Essentially, the definition is such that for two G-valued independent random variables $X \sim \mu$ and $Y \sim \nu$, $\mu * \nu$ is just the law of their product XY. For a compact group, the Haar measure μ_G is invariant under convolution, i.e. $\mu_G * \nu = \nu * \mu_G = \mu_G$ for all measures ν. It is easy to see that if μ and ν are distributions of linear operators, the moments of their convolution $\mu * \nu$ are given by their composition

$$M_\mu^k \circ M_\nu^k = M_{\mu * \nu}^k. \tag{B.1}$$

For many purposes it is useful to have a notion of distance between probability measures in order to quantify how much two probability distributions differ from another.

Definition B.10 (Total variation and Wasserstein distance)**.** Let $(S, \mathscr{A})$ be a mesurable space. For two measures μ, ν on S, we define

1. the *total variation distance*

$$d_{TV}(\mu,\nu) := \sup_{A\in\mathscr{A}} |\mu(A)-\nu(A)|.$$

2. If S is equipped with a metric d and $\mathscr{A}=\mathscr{B}(S)$, we can also define the *L^p-Wasserstein distance*:

$$\mathcal{W}_p(\mu,\nu) := \inf_{\Pi} \mathbb{E}_{\Pi}[d(x,y)^p]^{\frac{1}{p}}$$

where the infimum is taken over all measures on $S\times S$ with marginal distributions μ,ν,

Proposition B.11 (Kantorovich-Rubinstein duality, Theorem 5.10 in [83])**.** When two measures μ,ν on the metric space (S,d) have bounded support, the L^1-Wasserstein distance has the dual representation

$$\mathcal{W}_1(\mu,\nu) = \sup_{f\in \mathrm{Lip}_1(S)} \Big| \int f d\mu - \int f d\nu \Big|,$$

where the infimum is taken over all 1-Lipshitz functions from S to $\mathbb{R}$.

If the metric space (S,d) is bounded, we can estimate the L^1-Wasserstein distance against the total variation by

$$\mathcal{W}_1(\mu,\nu) \le \mathrm{diam}(S) d_{TV}(\mu,\nu)$$

where $\mathrm{diam}(S)<\infty$ is the diameter of S. This can be easily shown using the Kantorovich-Rubinstein duality.

Definition B.12. A probability (or Markov) kernel between two measurable spaces $(S_1,\mathscr{A}_1),(S_2,\mathscr{A}_2)$ is a map $\kappa : S_1\times\mathscr{A}_2\to\mathbb{R}$ such that

1. $\kappa(x_1,\cdot):\mathscr{A}_2\to\mathbb{R}$ defines a probability measure for all $x_1\in S_1$,
2. $\kappa(\cdot,A_2):S_1\to\mathbb{R}$ is measurable for all $A_2\in\mathscr{A}$.

For two probability kernels $(S_1,\mathscr{A}_1)\to(S_2,\mathscr{A}_2)$ and $\kappa_2:(S_2,\mathscr{A}_2)\to(S_3,\mathscr{A}_3)$, we define their composition by

$$(\kappa_2\circ\kappa_1)(x,A) = \int \kappa_2(y,A)\kappa_1(x,dy),$$

where we write $\kappa(x,dy)$ in order to indicate integration with repsect to the probability measure $\kappa(x,\cdot)$. Probability kernels allow us to push probability measures from one measurable space to another. With a probability kernel $\kappa:(S_1,\mathscr{A}_1)\to(S_2,\mathscr{A}_2)$ and a probability measure $\mathcal{P}$ on $(S_1,\mathscr{A}_1)$, we can construct a probability measure $\mathcal{P}^\kappa$ on $(S_2,\mathscr{A}_2)$ by

$$\mathcal{P}^\kappa(A) := \int_{S_1} \kappa(x,A) d\mathcal{P}(x)$$

for all $A\in\mathscr{A}_2$. A measurable map between measurable spaces $f:(\Omega_1,\mathscr{A}_1)\to(\Omega_2,\mathscr{A}_2)$, canonically defines a probability kernel between these spaces by

$$\kappa(\omega,A) = \begin{cases} 1, & f(\omega)\in A,\\ 0, & \text{otherwise.}\end{cases}$$

Then for any probability measure $\mathcal{P}$ on Ω_1, we have

$$\mathcal{P}^\kappa(A) = \int_{\Omega_1} \kappa(\omega, A) d\mathcal{P}(\omega) = \mathcal{P}(\{\omega : f(\omega) \in A\}) = \mathcal{P}^f(A).$$

This justifies the notation $\mathcal{P}^\kappa$ since it generalises the pushforward of measures along measurable maps.

Divergences

It is often useful to determine how much 'knowledge' two probability distributions have about each other. This can be measured for instance by the relative entropy, also known as the Kullback-Leibler divergence.

Definition B.13. Let μ, ν be probability measures on a measurable space $(S, \mathscr{A})$. We define the Kullback-Leibler divergence $\mathcal{D}_{KL}(\mu||\nu) \in [0, \infty]$ by

$$\mathcal{D}_{KL}(\mu||\nu) = \begin{cases} \mathbb{E}_\mu \left[\log\left(\frac{d\mu}{d\nu}\right)\right], \text{ if } \mu \text{ admits a density with respect to } \nu \\ \infty, \text{ otherwise.} \end{cases}$$

Note that we can rewrite $\mathcal{D}_{KL}(\mu||\nu) = \mathbb{E}_\mu \left[\log(\frac{d\mu}{d\nu})\right] = \mathbb{E}_\nu \left[\log\left(\frac{d\mu}{d\nu}\right)\frac{d\mu}{d\nu}\right]$ using the density. Since $x \mapsto x\log(x)$ is convex, it follows from Jensen's inequality that $\mathcal{D}_{KL}$ is nonnegative. The relative entropy has several useful properties, the most important ones are discussed in the following.

Lemma B.14. The Kullback-Leibler divergence is *convex*, i.e. for μ_1, μ_2, ν and $0 \leq \lambda \leq 1$:

$$\mathcal{D}_{KL}(\lambda\mu_1 + (1-\lambda)\mu_2||\nu) \leq \lambda\mathcal{D}_{KL}(\mu_1||\nu) + (1-\lambda)\mathcal{D}_{KL}(\mu_2||\nu) \tag{B.2}$$

Proof. Without loss of generality let μ_1, μ_2 be absolutely continuous with respect to ν (i.e. they have densities). Using the density of the convex combination $\lambda\mu_1 + (1-\lambda)\mu_2$ with respect to ν, we can express the relative entropy by

$$\mathcal{D}_{KL}\left(\lambda\mu_1 + (1-\lambda)\mu_2||\nu\right) = \mathbb{E}_\nu \left[\log\left(\lambda\frac{d\mu_1}{d\nu} + (1-\lambda)\frac{d\mu_2}{d\nu}\right)\left(\lambda\frac{d\mu_1}{d\nu} + (1-\lambda)\frac{d\mu_2}{d\nu}\right)\right].$$

Since the map $x \mapsto x\log(x)$ is convex, the claim follows. □

Proposition B.15. The relative entropy is decreasing under probabilistic operations. That is, for any pair of measurable spaces $(S_1, \mathscr{A}_1), (S_2, \mathscr{A}_2)$ and any probability kernel $\kappa : S_1 \times \mathscr{A}_2 \to \mathbb{R}$, we have

$$\mathcal{D}_{KL}(\mathcal{P}^\kappa||\mathcal{Q}^\kappa) \leq \mathcal{D}_{KL}(\mathcal{P}||\mathcal{Q}) \tag{B.3}$$

for all probability measures $\mathcal{P}, \mathcal{Q}$ on $(S_1, \mathscr{A}_1)$.

Proof. In the following we will freely write expectations as integrals or vice versa whenever this is opportune. Without loss of generality assume that $\mathcal{P}$ is absolutely continuous with respect to $\mathcal{Q}$. Let $\widetilde{\mathcal{P}}$ denote the measure on $S_1 \times S_2$ defined by

$$\widetilde{\mathcal{P}}(A \times B) = \int_A \int_B \kappa(x, dy) d\mathcal{P}(x),$$

analogously for $\mathcal{Q}$. Let π_1, π_2 denote the respective projections. We observe that

$$\widetilde{\mathcal{P}}(A \times B) = \int_{A\times B} \kappa(x, dy) d\mathcal{P}(x) = \int_{A\times B} \frac{d\mathcal{P}}{d\mathcal{Q}}(x)\kappa(x, dy)d\mathcal{Q}(x) = \int_{A\times B} \frac{d\mathcal{P}}{d\mathcal{Q}}(x)d\widetilde{\mathcal{Q}}(x, y).$$

So the density of $\widetilde{\mathcal{P}}$ with respect to $\widetilde{\mathcal{Q}}$ exists and is given as the pullback $\pi_1^*\left(\frac{d\mathcal{P}}{d\mathcal{Q}}\right)$ along the projection onto the first component. Its conditional expectation with respect to the $\sigma-$subalgebra $\{\emptyset, S_1\} \times \mathscr{A}_2 \subset \mathscr{A}_1 \otimes \mathscr{A}_2$ which is induced by the projection π_2 satisfies

$$\begin{aligned}&\int_B \mathbb{E}_{\widetilde{\mathcal{Q}}}\left[\frac{d\widetilde{\mathcal{P}}}{d\widetilde{\mathcal{Q}}}\middle|\pi_2\right] d\mathcal{Q}^\kappa(y) = \int_{S_1\times B} \mathbb{E}_{\widetilde{\mathcal{Q}}}\left[\frac{d\widetilde{\mathcal{P}}}{d\widetilde{\mathcal{Q}}}\middle|\pi_2\right] d\widetilde{\mathcal{Q}}(x, y)\\ =&\int_{S_1\times B} \frac{d\widetilde{\mathcal{P}}}{d\widetilde{\mathcal{Q}}} d\widetilde{\mathcal{Q}} = \int_{S_1\times B} d\widetilde{\mathcal{P}} = \mathcal{P}^\kappa(B)\end{aligned}$$

for all $B \in \mathscr{A}_2$. Therefore the density of $\mathcal{P}^\kappa$ with respect to $\mathcal{Q}^\kappa$ exists and it is given by

$$\frac{d\mathcal{P}^\kappa}{d\mathcal{Q}^\kappa} = \mathbb{E}_{\widetilde{\mathcal{Q}}}\left[\frac{d\widetilde{\mathcal{P}}}{d\widetilde{\mathcal{Q}}}\middle|\pi_2\right]. \tag{B.4}$$

In particular, we have $\mathcal{D}_{KL}(\mathcal{P}^\kappa||\mathcal{Q}^\kappa) < \infty$. We can rewrite the Kullback-Leibler divergence as

$$\mathcal{D}_{KL}(\mathcal{P}^\kappa||\mathcal{Q}^\kappa) = \int_{S_2} \log\left(\frac{d\mathcal{P}^\kappa}{d\mathcal{Q}^\kappa}\right) d\mathcal{P}^\kappa = \int_{S_2} \log\left(\mathbb{E}_{\widetilde{\mathcal{Q}}}\left[\frac{d\widetilde{\mathcal{P}}}{d\widetilde{\mathcal{Q}}}\middle|\pi_2\right]\right) \mathbb{E}_{\widetilde{\mathcal{Q}}}\left[\frac{d\widetilde{\mathcal{P}}}{d\widetilde{\mathcal{Q}}}\middle|\pi_2\right] d\mathcal{Q}^\kappa.$$

Next, we apply Jensen's inequality to the convex function $x \mapsto x\log(x)$

$$\leq \int_{S_2} \mathbb{E}_{\widetilde{\mathcal{Q}}}\left[\log\left(\frac{d\widetilde{\mathcal{P}}}{d\widetilde{\mathcal{Q}}}\right)\frac{d\widetilde{\mathcal{P}}}{d\widetilde{\mathcal{Q}}}\middle|\pi_2\right] d\mathcal{Q}^\kappa = \int_{S_1\times S_2} \mathbb{E}_{\widetilde{\mathcal{Q}}}\left[\log\left(\frac{d\widetilde{\mathcal{P}}}{d\widetilde{\mathcal{Q}}}\right)\frac{d\widetilde{\mathcal{P}}}{d\widetilde{\mathcal{Q}}}\middle|\pi_2\right] d\widetilde{\mathcal{Q}}.$$

With the law of total expectation we can conclude

$$\begin{aligned}&= \int_{S_1\times S_2} \log\left(\frac{d\widetilde{\mathcal{P}}}{d\widetilde{\mathcal{Q}}}\right) d\widetilde{\mathcal{P}} = \int_{S_1\times S_2} \log\left(\frac{d\mathcal{P}}{d\mathcal{Q}}(x)\right)\kappa(x, dy)d\mathcal{P}(x)\\ &= \int_{S_1} \log\left(\frac{d\mathcal{P}}{d\mathcal{Q}}\right) d\mathcal{P} = \mathcal{D}_{KL}(\mathcal{P}||\mathcal{Q}).\end{aligned}$$

□

For probability measures μ, ν on some (semi)group, the convolution of measures with another measure η is a special instance of a Markov kernel. Accordingly, the above estimate (B.5) reads

$$\mathcal{D}_{KL}(\mu * \eta||\nu * \eta) \leq \mathcal{D}_{KL}(\mu||\nu) \tag{B.5}$$

in this case.

Proposition B.16 (Pinsker's inequality, Lemma 2.5 in [84])**.** Given two probability measures μ, ν on a measurable space $(S, \mathscr{A})$, we can estimate their total variation distance against the

Kullback-Leibler divergence by

$$d_{TV}(\mu,\nu) \leq \sqrt{\frac{1}{2}\mathcal{D}_{KL}(\mu||\nu)}. \tag{B.6}$$

B.3 Stochastic Analysis

Definition B.17 (Stochastic process)**.** A stochastic process on a probability space $(\Omega, \mathcal{P}, \mathscr{F})$ with values in a measurable space $(S, \mathscr{A})$ is a collection of random variables $(X_i)_{i\in I}$ with $X_i : \Omega \to S$ for some index space $I \subset \mathbb{R}$.

Remark B.18. Natural choices for the index set I are $\mathbb{N}$ or $\mathbb{Z}$, but for us this will be almost exclusively $I = [0, \infty)$. Whenever we are given a stochastic process $X : \Omega \times I \to S$ on a probability space $(\Omega, \mathcal{P}, \mathscr{F})$, we usually omit the dependence of X on $\omega \in \Omega$, and in most cases also the indication of the index set. X_t then refers to the random variable $\omega \mapsto X(t, \omega)$ and the sample paths of X are the maps $X(\omega) : \mathbb{R}_{\geq 0} \to S$. In order to construct stochastic processes through probability distributions of random variables X_i which meet some compatibility criteria, we can rely on Kolmogorovs Extension Theorem which is explained in Chapter 14.3 of [81].

Modelling a stochastic process on a probability space $(\Omega, \mathcal{P}, \mathscr{F})$ adds additional structure to the model which motivates the notion of filtered probability spaces.

Definition B.19.

1. A *filtration* of a probability space $(\Omega, \mathcal{P}, \mathscr{F})$ is a family of σ-algebras $(\mathscr{F}_t)_{t\in I\subset\mathbb{R}_{\geq 0}}$ such that
$$\mathscr{F}_\infty := \sigma\left(\bigcup_{t\geq 0} \mathscr{F}_t\right) = \mathscr{F} \text{ and } \mathscr{F}_s \subset \mathscr{F}_t \text{ for } s \leq t.$$
2. A stochastic process X on a filtered probability space $(\Omega, \mathcal{P}, (\mathscr{F}_t)_t)$ is called *adapted* if X_t is $\mathscr{F}_t$-measurable for all $t \geq 0$.

The condition on being adapted means that by collecting all information available up to time $t \geq 0$, we can distinguish between all events in $\mathscr{F}_t$. For a stochastic process X, we will always understand the filtration to be generated by X, i.e. given by $\mathscr{F}_t := \sigma((X_s)_{s\leq t})$, unless specified otherwise. By saying that a stochastic process X is adapted to another process Y, we mean that X is adapted to the filtration generated by Y in this fashion.

Often one encounters processes which are not continuous in time, but have jumps. Because such jumps happen suddenly and cannot be anticipated, the sample paths should meet the requirements of the next definition.

Definition B.20. A function $f : \mathbb{R} \to S$ into some metric space S is called *càdlàg* (continue à droite, limites à gouche) if its left limits exist and it is right continuous, i.e.

$$\lim_{s\to t_-} f(s) \in \mathbb{R} \quad \text{and} \quad \lim_{s\to t_+} f(s) = f(t).$$

hold or all $t \geq 0$.

An important class of stochastic processes are Markov processes. These are processes which have no memory in the sense that their future development only depends on the current value, and not on their entire history until to this point.

Definition B.21 (Markov process, Def. 17.1,Thm. 17.8 in [81])**.** A stochastic process $(X_i)_{i\in I}$ on a probability space $(\Omega, \mathcal{P}, \mathscr{F})$ with values in a measurable state space S is called a *Markov process* if one of the two equivalent conditions is satisfied:

1. We have $\mathbb{E}[f(X_j)|\mathscr{F}_i] = \mathbb{E}[f(X_j)|\sigma(X_i)]$ for all bounded measurable $f : S \to \mathbb{R}$ and $j \geq i$, where $(\mathscr{F}_i)_{i\in I}$ denotes the filtration generated by X.
2. There exists a family $(\pi_{j,i})_{i\leq j\in I}$ of probability kernels subject to the *Chapman-Kolmogorov equations* that is $\pi_{i,j} \circ \pi_{j,k} = \pi_{i,k}$ for all $i \geq j \geq k$, such that $\mathcal{P}^{X_i} = \pi_{i,j}(\mathcal{P}^{X_j})$ holds for all $i \geq j$.

We use the notation π instead of κ to indicate probability kernels that belong to a Markov process.

Brownian motion and white noise

The central type of process around which stochastic analysis evolves are Brownian motions:

Definition B.22. A real valued stochastic process $(B_t)_{t\geq 0}$ is called *Brownian motion* (sometimes also called Wiener process) if

1. the increments $B_{t_n} - B_{t_{n-1}}, \ldots, B_{t_1} - B_{t_0}$ are independent for all $t_n > \cdots > t_0$ and stationary, i.e. $B_t - B_s \sim B_{t-s}$ for all $t \geq s$,
2. B_t is Gaussian with mean zero and variance $\sqrt{t}$, i.e. $B_t \sim \mathcal{N}(0, t)$ for $t > 0$,
3. $B_0 = 0$.

Brownian motion is almost surely nowhere differentiable, as one can easily tell by looking at its increments

$$\frac{B_t - B_s}{t-s} \sim \frac{B_{t-s}}{t-s} \sim \frac{B_1}{\sqrt{t-s}}$$

However it is possible to define a generalised derivative in terms of tempered distributions.

Theorem B.23 (Bochner-Minlos, Thm 2.1.1 in [85])**.** Let $\mathscr{S}\left(\mathbb{R}^d\right)$ be the Schwartz space, $\mathscr{S}'\left(\mathbb{R}^d\right)$ the space of tempered distributions and let $\langle\cdot,\cdot\rangle$ denote the pairing between these spaces. Then there exists a unique probability measure μ on the Borel sets $\mathscr{B}\left(\mathscr{S}'\right)$, called *white noise*, such that we have

$$\mathbb{E}\left[e^{i\langle\omega,\phi\rangle}\right] = \int_{\mathscr{S}'(\mathbb{R}^d)} e^{i\langle\eta,\phi\rangle} d\mu\left(\eta\right) = e^{-\frac{1}{2}\|\phi\|_{L^2}^2}.$$

for any $\phi \in \mathscr{S}\left(\mathbb{R}^d\right)$ and $\omega \sim \mu$.

How is this linked to Brownian motion? The white noise measure μ is the unique measure on $\mathscr{S}'$ such that, for $\omega \sim \mu$ and any sequence of compactly supported functions $\phi_n \in C_c^\infty(\mathbb{R})$ approximating the characteristic funtion of the interval $[0, t]$, the limit

$$B_t := \lim_{n\to\infty} \omega(\phi_n) \left(= \int_0^t \omega_s ds, \text{ formally}\right)$$

defines a Brownian motion. One more important characteristic feature of white noise is that its autocorrelation in the distributional sense corresponds to the Dirac delta distribution. For ω distributed according to white noise, we have

$$\mathbb{E}[\omega \otimes \omega] = \delta_{diag},$$

where $\delta_{diag} \in \mathscr{S}'(\mathbb{R} \otimes \mathbb{R})$ is the Dirac delta supported on the diagonal. Of course one could extend the discussion about the correspondence between white noise and Brownian motion further at this point, however it is not strictly necessary for us, so we shall leave it at that. More details on this can be found for instance in [85].

Martingales

Martingales are one of the most fundamental objects in modern probability theory and used to model 'fair gambling' processes. We only introduce as much of the theory as is immediately needed to introduce stochastic integration. This in principle includes a lot of powerful estimates and convergence results, which can be derived for martingales. However these are more needed on a technical level and not strictly necessary to get a general idea of the topic which is what we are aiming at. Therefore, we will not included these into our summary here and hint at Chapter 11 in [81] instead.

Definition B.24 (Martingale)**.** A stochastic process $(M_t)_t$ on a filtered probability space $(\Omega, \mathcal{P}, (\mathscr{F}_t)_t)$ is called a *martingale* if

1. M is adapted to the filtration $(\mathscr{F}_t)_t$,
2. $\mathbb{E}[|M_t|] \leq \infty$ for all $t \geq 0$,
3. $\mathbb{E}[M_t|\mathscr{F}_s] = M_s$ for all $t \geq s \geq 0$.

In particular martingales have the property that their expectation remains constant. One could also stop processes after some random time, leading to the concept of stopping times.

Definition B.25. A map $T : \Omega \to [0, \infty]$ on a filtered probability space $(\Omega, \mathcal{P}, (\mathscr{F}_t)_t)$ is called a $\mathscr{F}_t$-stopping time if $\{T \leq t\} \in \mathscr{F}_t$ for all $t \in [0, \infty]$.

Doob's optional sampling theorem shows that martingale properties are preserved under stopping.

Proposition B.26 (Doob's optional sampling theorem, Thm. 3.22 in [46])**.** Let X be a continuous martingale and T a stopping time. Then $X_t^T := X_{T \wedge t}$ is again a martingale, where we denote $T \wedge t := \min\{T, t\}$.

There are two important classes of processes which are not martingales, but are closely related.

Definition B.27.

1. A stochastic process X is called a *local martingale* if there exists an increasing sequence of stopping times T_n such that $X_t^n = X_{t \wedge T_n}$ is a martingale for every n and $T_n \to \infty$ almost surely.
2. A stochastic process X is called a *semimartinagle* if it admits a decomposition $X_t = X_0 + M_t + V_t$, where M is a martingale and V is a process with bounded first variation (see below) such that $M_0 = V_0 = 0$.

For an adapted process X and a partition $\Pi = \{t_0, \ldots, t_n\}$ of the interval $[0, t]$ we define the p-variation ($p > 0$) of X with respect to Π by

$$V_{X,t}^{(p)}(\Pi) := \sum_{k=1}^{n} |M_{t_k} - M_{t_{k-1}}|^p.$$

For a partition we write $\|\Pi\| := \max_{1\leq j\leq n} |t_j - t_{j-1}|$ for its mesh size and $|\Pi| = n$ for its size.

Theorem B.28 (Thm 5.8, Def. 5.3 in [46])**.** Let $X = X_0 + V + M$ a semimartimgale as above such that M is square integrable. Then we have the following:

1. The limit
$$[X]_t := \lim_{\|\Pi\|\to 0} V^{(2)}_{X,t}(\Pi)$$
exists almost surely and, even stronger, in probability[1]. The process $([X]_t)_t$ is called the *quadratic variation* of X.

2. For a square integrable martingale M, the quadratic variation $([M]_t)_{t\geq 0}$ is the unique increasing process such that $M_t^2 - [M]_t$ is again a martingale (up to probabalistic indistiguishability).

It is not difficult to show directly that the quadratic variation of a Brownian motion B is given by $[B]_t = t$. By a result of Lévy, a square-integrable local martingale is a Brownian motion if and only if its quadratic variation is of that form, see p. 157 in [46]. This is a very well-behaved example of a quadratic variation. We want to emphasize that in general the outcome is not necessarily a deterministic function.

Definition B.29. Let X, Y be semi-martingales with square-integrable martingale parts. We define their *cross variation* to be the process
$$[X, Y]_t := \frac{1}{4}\left([X + Y]_t - [X - Y]_t\right). \tag{B.7}$$

It is immediate that we can express the cross variation via the limit
$$[X, Y]_t = \lim_{\|\Pi\|\to 0} \sum_{k=0}^{|\Pi|} (X_{t_{k+1}} - X_{t_k})(Y_{t_{k+1}} - Y_{t_k}). \tag{B.8}$$

Remark B.30. Clearly the cross variation of a process X with itself gives back its quadractic variation $[X, X] = [X]$. Of course $[X, Y]$ is not increasing anymore as opposed to the quadratic variation. However, we can write it as the difference of two increasing processes, and therefore $[X, Y]$ still has bounded first variation. It is not difficult to show that the cross variation (B.8) vanishes if the processes are independent or if one of the processes has bounded first variation. With polarisation, one can also easily show that $MN - [M, N]$ is again a martingale if M and N are martingales. When dealing with processes X, Y that take values in $\mathbb{R}^n$ we use the shorthand notation $[X, Y] = ([X_1, Y_1], \ldots, [X_n, Y_n])$.

[1] $X_n \to Y$ in probability if for every $\varepsilon, \delta > 0$ there is N such that $\mathcal{P}(|X_n - Y| > \delta) < \varepsilon$ for all $n \geq N$.

C. REMARKS ON SIMULATIONS

For our examples we want to simulate solutions to the random differential equation

$$\frac{d}{dt}U_t = -\Big(iH_0 + i\sum_j c_j(t)H_j^c\Big)U_t, quadU_0 = 1_G \tag{C.1}$$

where the c_j are piecewise constant, and normally distributed $\sim \mathcal{N}(0, \Delta T^{-1})$ for a time step $\Delta T > 0$. Following the results of Stroock and Varadhan of Thm.4.21 this corresponds to an approximation to the solution of the Stratonovich-SDE

$$dU_t = -\Big(iH_0 dt + i\sum_j H_j^c \circ dB_t^j\Big)U_t, \quad U_0 = 1_G \tag{C.2}$$

for some independent Brownian motions $B^1, \ldots, B^m$. Since the Hamiltonian in (C.1) is time independent on the interval $[0, \Delta T]$ we can solve this random differential equation locally by ordinary exponentiation. By multiplication of these increments we recover the expression (5.5) from the introduction of Chapter 5 (up to multiplication order and signs). In order to ensure that the evolution is not affected by roundoffs too much, we simulate the distribution at $t = \Delta T$ once for 10^6 samples and compute the exponential with high precision. We will then use the uniform distribution over this collection as a substitute for the original one. In order to compute the trace norm of the k-th moments as memory-efficient as possible, we rely on Remark 3.7 and determine the corresponding frame potential of the sampled distribution. In the programming language Julia [86], the result is the following.

```
using LinearAlgebra;
using Random;

function Simulation()
  ################ Auxiliary Functions ####################
  function exp(A)
    x = norm(A)
    ind = 0
    n = size(A)[1]
    I_n = (1+0im)*Matrix(I, n, n)
    while 2^ind < round(x)
      ind += 1
    end
    aux = I_n
    e = I_n
    A_new = A/(2^ind)
    for i = 1:50
      aux = A_new*aux/i
      e = e + aux
    end
    for i = 1:ind
      e = e*e
```

```
23      end
24      return e
25    end
26    ################ Initialising ##########################
27    qubits =                  # number of qubits
28    drift =                   # define a drift Hamiltonian
29    controls =                # specify a list of control Hamiltonians
30    samples =                 # Number of samples
31    moment =                  # specify tensor power of \Phi^{1,1}
32    rate =                    # sampling rate of determining frame potential
33    time =                    # maximum evolution time (1 timestep = pi)
34    resolution =              # number of iterations per time step (pi/Delta_t)
35    ################ Actual Programme ######################
36    N = resolution*time
37    dim = 2^qubits
38    delta_t = pi*time/N
39    Distribution = Vector{Matrix{ComplexF64}}(undef,100000)
40    Evolution = Vector{Matrix{ComplexF64}}(undef,Max)
41    rng = MersenneTwister(1234)
42    for j = 1:100000
43      Increment = (1+0im)*zeros(ComplexF64, dim,dim)
44      for i = 1:(size(controls)[1])
45        rdm = randn(rng, Float64)
46        Increment = Increment+rdm*controls[i]
47      end
48      Distribution[j] = exp(-1im*delta_t*drift - 1im*sqrt(delta_t)*Increment)
49    end
50    for j = 1:samples
51      Evolution[j] = Matrix{ComplexF64}(I,dim,dim)
52    end
53    results = Vector{Float64}()
54    append!(results,2^(2*qubits*moment))
55    for t = 1:N
56      for j = 1:samples
57        m = rand(1:100000)
58        Evolution[j] = Distribution[m]*Evolution[j]
59      end
60      if round(rate*t)%resolution == 0
61        value = 0
62        for i = 1:samples
63          for j = 1:samples
64            aux = abs(tr(adjoint(Evolution[i]) * Evolution[j]))^(2*moment)
65            value += aux/(samples^2)
66          end
67        end
68        append!(results,value)
69      end
70    end
71    return results
72  end
73    #output: list containing the squared trace norm of the second
74    #        moments of the distribution at times (no. of entry *pi)/rate
```

D. NOTATIONS AND SYMBOLS

$(\cdot)^\dagger$	hermitian conjugate of a matrix/ adjoint of an operator
$\odot dM$	Stratonovich stochastic increment
$*$	convolution
$\bar{(\cdot)}$	complex conjugate
$\lhd$	ideal
$\mathscr{A}$	σ-algebra
$\mathcal{A}_t(x_0)$	attainable ste by a control system at time t starting at x_0
$(B_t)_t$	Wiener process/Brownian motion
$\mathscr{B}(S)$	Borel σ-algebra of a toplogical space S
$C^k(M)$	$k-$times continuously differentiable functions on M
C_c	compactly supported functions
$D_x f$	differential of f at x as a linear map
$\mathcal{D}$	the space of test-functions $(= C_c^\infty)$
$\mathcal{D}'$	space of distributions
Dom	domain of a unbounded operator
$\mathcal{D}_{KL}$	Kullback-Leibler divergence
d_{TV}	total variation distance
$\mathbb{E}[Z]$	expectation of a random variable
$\mathbb{E}[Z\|\mathcal{G}]$	conditional expectation of Z with respect to $G \subset \mathscr{F}$
$\mathrm{End}(V)$	endomorphisms on a finite-dimesnional vector space V
$\mathscr{F}$	σ-algebra of a probability space
$(\mathscr{F}_t)_t$	filtration of a sigma-algebra $\mathscr{F}$
$(g_t)_{t\geq 0}$	Lévy process on a Lie group
$Gl(V)$	group of invertible linear operators on finite-dimensional vetor space
G	Lie group
G_0	controlled Lie group
$\mathfrak{g}$	Lie algebra (mostly of the Lie group G)
$\mathfrak{g}, \mathfrak{h}$	Lie algebras
$\mathfrak{g}_0$	control algebra
$\mathcal{H}$	Hilbert space
$\mathfrak{I}_0$	zero-time ideal
$\mathcal{K}$	Killing form
κ	probability kernel
Λ	quantum channel
$L^p(\mu)$	$p-$integrable functions with respect to μ
$L(\mathcal{H})$	linear operators on a Hilbert space
$\mathcal{L}$	infinitesimal generator of Feller processes
$\mathcal{L}_\phi$	infinitesimal generator of moment-semigroup
$\mathcal{M}(\mathcal{X}, \mathcal{Y})$	measureable functions between two measureable spaces $\mathcal{X}, \mathcal{Y}$
$\mathcal{M}^1(S)$	set of probability measures on measureable space $(S, \mathscr{A})$

M_μ^Φ	moment of a measure μ on a Lie group with respect to representation Φ
μ, ν, η	measures on a measureable space
$\{\mu_t\}_t$	convolution semigroup associated to a Lévy process $(g_t)_t$ on Lie group
μ_G	the Haar measure on a compact group G
M, N	martingales
$[M]$,$[M, N]$	quadratic and crossvariation processes respectively
Ω	space of outcomes of a probability space
ω	outcomes on a probability space
Φ, Ψ	Lie group representations
$\Phi^{k,k}$	representation on subgroups of $\mathrm{SU}(d)$ given by $U \mapsto U^{\otimes k} \otimes \bar{U}^{\otimes k}$
ϕ	Lie algebra representations, mostly corresponding to Φ
$\phi^{k,k}$	Lie algebra representation corresponding to $\Phi^{k,k}$
ψ	quantum state
$\{P_t\}_t$	probability operator semigroup of Markov process
$\mathcal{P}$	probability measure
p_t	density of the Law μ_t of a Lévy process with respect to invariant measure
$\pi_{t,s}$	transition kernels of a Markov process
$q(\cdot, \cdot)$	Riemannian metric
$\Re, \Im$	real/imaginary part
ρ	mixed state density operator of a quantum system
$\mathscr{S}(\mathbb{R}^d)$	Schwartz space
$\mathscr{S}'(\mathbb{R}^d)$	tempered distributions
$\sigma_x, \sigma_y, \sigma_z$	Pauli Matrices
$\mathfrak{su}_d, \mathfrak{so}_d$	special unitary, special orthogonal and compact symplectic Lie algebras
$\int Z dM$	Itô integral
$\int Z \circ dM$	Stratonovich integral
$T_x M$	tangent space of M at x
Tr	trace
$U(\mathfrak{g})$	universal envelopping algebra
U_t	unitary evolution of control system
$VF(M)$	Lie algebra of smooth vector fields on a smooth manifold M
$V^G, V^\mathfrak{g}$	invariant subspaces of a Lie group/algebra representation on a vector space V
V, W	general vector fields
$\mathcal{W}_1$	L^1-Wasserstein/Kantorovich-Rubinstin distance
χ_A	characteristic function of a set A
ξ	random variable with values $\mathscr{S}'$ distributed according to white noise
$(X\|Y_1, \ldots, Y_k)$	control system
x, y, z	usually Elements of a Lie algebra
X, Y, Z	stochastic processes *or* vector fields,
	on a Lie group X, Y, Z are always the *left-invariant* vector fields to $x, y, z \in \mathfrak{g}$
$\mathcal{Y}$	controlled distribition of a control system
$\mathfrak{z}(\mathfrak{g})$	center of the Lie algebra $\mathfrak{g}$
$\mathcal{Z}(G)$	center of a Lie group

BIBLIOGRAPHY

[1] R. Feynman. “Simulating Physics with Computers”. *Int. J. Theor. Phys.* 21 (1982), 467–488.

[2] P. W. Shor. “Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer”. *SIAM J. Comput.* 26 (1997), 1484–1509.

[3] A.G. Fowler; M. Mariantoni; J.M. Martinis; A.N. Cleland. “Surface codes: Towards practical large-scale quantum computation”. *Phys. Rev. A* 86 (2012).

[4] I.M. Georgescu; S. Ashhab; F. Nori. “Quantum simulation”. *Rev. Mod. Phys.* 86 (2014), 153–185.

[5] E. Farhi; J. Goldstone; S. Gutmann. *A Quantum Approximate Optimization Algorithm*. https://arxiv.org/abs/1411.4028v1. 2014.

[6] M.P. Harrigan; K.J. Sung; M. Neeley; K.J. Satzinger; F. Arute; K. Arya; J. Atalaya; J.C. Bardin; R. Barends; S. Boixo et al. “Quantum approximate optimization of non-planar graph problems on a planar superconducting processor”. *Nat. Phys.* 17 (2021), 332–336.

[7] L. Zhou; S.-T. Wang; S. Choi; H. Pichler; M.D. Lukin. “Quantum Approximate Optimization Algorithm: Performance, Mechanism, and Implementation on Near-Term Devices”. *Phys. Rev. X* 10 (2020), 021067.

[8] M. Cerezo; A. Arrasmith; R. Babbush; S.C. Benjamin; S. Endo; K. Fujii; J.R. McClean; K. Mitarai; X. Yuan; L. Cincio; P.J. Coles. “Variational Quantum Algorithms”. *Nat. Rev. Phys.* 3 (2021), 625–644.

[9] J. Tilly; H. Chen; S. Cao; D. Picozzi; K. Setia; Y. Li; E. Grant; L. Wossnig; I. Rungger; G.H. Booth; J. Tennyson. *The Variational Quantum Eigensolver: a review of methods and best practices*. https://arxiv.org/abs/1602.04177. 2021.

[10] Z. Wang; S. Hadfield; Z. Jiang; E.G. Rieffel. “Quantum approximate optimization algorithm for MaxCut: A fermionic view”. *Phys. Rev. A* 97 (2018), 022304.

[11] C. Dankert; R. Cleve; J. Emerson; E. Livine. “Exact and approximate unitary 2-designs and their application to fidelity estimation”. *Phys. Rev. A* 80 (2009), 012304.

[12] J. Emerson; R. Alicki; K.Zyczkowski. “Scalable noise estimation with random unitary operators”. *J. Opt. B* 7 (2005), S347.

[13] E. Derbyshire; J. Yago Malo; A.J. Daley; E. Kashefi; P. Wallden. “Randomized benchmarking in the analogue setting”. *Quantum Sci. Technol.* 5 (2020), 049601.

[14] E. Magesan; J.M. Gambetta; J. Emerson. “Scalable and Robust Randomized Benchmarking of Quantum Processes”. *Phys. Rev. Lett.* 106 (2011), 180504.

[15] J. Helsen; X. Xue; L.M.K. Vandersypen; S. Wehner. “A new class of efficient randomized benchmarking protocols”. *npj Quantum Inf.* 5.71 (2019).

[16] Z. Webb. "The Clifford group forms a 3-design". *Quantum Inf. Comput.* 16 (2016), 1379–1400.

[17] A. W. Harrow; R. A. Low. "Random quantum circuits are approximate 2-designs". *Comm. Math. Phys.* 291 (2009), 257–302.

[18] E. Onorati; O.Buerschaper; M. Kliesch; W. Brown; A.H. Werner; J.Eisert. "Mixing properties of stochastic quantum Hamiltonians". *Comm. Math. Phys.* 355 (2017), 905–947.

[19] L. Banchi; D. Burgarth; M. J. Kastoryano. "Driven Quantum Dynamics - Will it blend?" *Phys.Rev. X* 7 (2017), 041015.

[20] D. Bakry; I. Gentil; M. Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. Springer, 2014.

[21] F. Baudoin. *Wasserstein contraction properties for hypoelliptic diffusions*. https://arxiv.org/abs/1602.04177. 2016.

[22] F. Baudoin. *Sub-Laplacians and hypoelliptic operators on totally geodesic Riemannian foliations*. https://arxiv.org/abs/1410.3268. 2014.

[23] F. Baudoin; N.Garofalo. "Curvature-dimension inequalities and Ricci lower bounds for sub-Riemannian manifolds with transverse symmetries". *J. Eur. Math. Soc.* 19 (2017), 151–219.

[24] C. Villani. *Hypocoercivity*. Memoirs of the America Mathematical society Vol. 202. American Mathematical Society, 2009.

[25] D. Bakry; M. Émery. "Diffusions hypercontractives". *Séminaire de probabilités de Strasbourg* 19 (1985), 177–206.

[26] R. Zeier; Z. Zimborás. "On squares of representations of compact Lie algebras". *J. Math. Phys.* 56 (2015), 081702.

[27] S. de Léséleuc; V. Lienhard; P. Scholl; D. Barredo; S. Weber; N. Lamg; H.P. Büchler; T. Lahaya; A. Browaeys. "Observation of a symmetry-protected topological phase of interacting bosons with Rydberg atoms". *Science* 365 (2019), 775–780.

[28] W. Bosma; J. Cannon; C. Playoust. "The Magma algebra system. I. The user language". *J. Symbolic Comput.* 24 (1997), 235–265.

[29] D. Gross; K. Audenaert; J. Eisert. "Evenly distributed unitaries: on the structure of unitary designs". *J. Math. Phys* 48 (2007), 052104.

[30] R.A. Low. "Pseudo-randomness and Learning in Quantum Computation". PhD thesis. University of Bristol, 2009.

[31] M.A. Nielsen; I.L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2016.

[32] G. Nebe; E.M. Rains; N.J.A. Sloane. "The Invariants of the Clifford Groups". *Des. Codes, Cryptogr* (2001), 99–122.

[33] J.L. Troutman. *Variational Claculus and Optimal Control*. Springer, 1996.

[34] V. Jurdjevic; H.J. Sussmann. "Control Systems on Lie Groups". *J. Differ. Equ.* 12 (1972), 313–329.

[35] G. Dirr; U. Helmke. "Lie Theory for Quantum Control". *GAMM Mitteilungen.* 31 (2008), 59–93.

[36] V. Jurdjevic. *Geometric Control Theory*. Cambridge Univ. Press, 1997.

[37] D. Elliott. *Bilinear Control Systems - Matrices in Action*. Springer, 2009.

[38] V. Jurdjevic; H.J. Sussmann. "Controllability of non-Linear Systems". *J. Differ. Equ.* 12 (1972), 95–116.

[39] T. H. Gronwall. "Note on the Derivatives with Respect to a Parameter of the Solutions of a System of Differential Equations". *Annals of Mathematics* 20 (1919), 292–296.

[40] P. Jouan. "Finite-time and Exact Controllability on Compact Manifolds". *J. Math. Sci* 177 (2011), 402.

[41] R. Zeier; T. Schulte-Herbrüggen. "Symmetry Principles in Quantum Systems Theory". *J. Math. Phys.* 52 (2011), 113510.

[42] J.J. Seidel. "Spherical Designs and Tensors". *Progress in Algebraic Combinatorics*. 24 vols. Mathematical Society of Japan, 1996.

[43] C. Dankert. "Efficient Simulation of Random Quantum States and Operators". https://arxiv.org/abs/quant-ph/0512217. MA thesis. University of Waterloo, 2005.

[44] F.G.S.L. Brandão; P. Ćwikliński; M. Horodecki; P. Horodecki; J.K. Korbicz; M. Mozrzymas. "Convergence to equilibrium under a random Hamiltonian". *Phys. Rev. E* 86 (2012), 031101.

[45] F.Hiai; D. Petz. *The Semicircle Law, Free Random Variables and Entropy*. American Mathematical Society, 2000.

[46] I. Karatzas; S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 1991.

[47] D.W. Stroock. *Markov Processes from K. Itôs Perspective*. Princeton University Press, 2003.

[48] K. Itô. "Stochastic Integral". *Proc. Imp. Acad.* 20.8 (1944), 519–524.

[49] R. Stratonovich. "A new Representation for Stochastic Integrals and Equations". *SIAM J. Control Optim.* 4 (1966), 362–371.

[50] A. Bovier. "Introduction to Stochastic Analysis". https://wt.iam.uni-bonn.de/bovier/lecture-notes.

[51] K. Itô. "Stochastic Differentials". *Appl. Math. Opt.* 1 (1974), 374–381.

[52] D.W. Stroock. *Elements of Stochastic Calculus and Analysis*. Springer, 2018.

[53] H. Kunita; S. Watanabe. "On square integrable martingales". *Nagoya Math. J.* 30 (1967), 209–245.

[54] K. Itô. "Extension of stochastic Integrals". *Proceedings of the International Symposium on Stochastic Differential Equations*. Res. lnst. Math. Sci., Kyoto Univ. 1976, pp. 95–109.

[55] D.W. Stroock; S.R.S. Varadhan. "On the support of diffusion processes with applications to the strong maximum principle". *Berkeley Smyp. on Math. Statist.and Prob.* 6.3 (1972), 333–359.

[56] H.J. Sussmann. "On the gap between ordinary deterministic and stochastic differential equations". *Ann. Prob.* 6 (1978), 19–41.

[57] P.E. Kloeden; E. Platen. *Numerical Solutions of Stochastic differential equations*. Springer, 1999.

[58] P. Aniello; A. Kossakowski; G. Marmo; F. Ventriglia. "Brownian motion on Lie groups and open Quantum systems". *J. Phys. A.* 43 (2010), 265301.

[59] Ming Liao. *Lévy processes in Lie groups*. Cambridge University Press, 2004.

[60] A. Eberle. *Markov Processes*. Lecture Notes. https://uni-bonn.sciebo.de/s/kzTUFff5FrWGAay. 2021.

[61] T.M. Liggett. *Continuous Time Markov Processes - An Introduction*. American Mathematical Society, 2010.

[62] W. Rudin. *Functional Analysis*. McGraw-Hill, 1991.

[63] J.B. Conway. *A Course in Functional Analysiws*. Springer, 1985.

[64] G.A. Hunt. "Semigroups of Measures on Lie Groups". *Trans. Am. Math. Soc.* 81.2 (1956), 264–293.

[65] K.-J. Engel; R. Nagel. *A short course on operator semigroups*. Springer, 2006.

[66] K. Yosida. *Functional Analysis*. Springer, 1968.

[67] L. Hörmander. "Hypoelliptic Second Order Differential Equations". *Acta Math.* 119 (1967), 147–171.

[68] D. Nualart. *The Malliavin Calculus and Related Topics*. Springer, 2006.

[69] F. Baudoin. *An Introduction to the Geometry of Stochastic Flows*. Imperial College Press, 2004.

[70] P. Harremoës. "Maximum Entropy on Compact Groups". *Entropy* 11 (2009), 222–237.

[71] K.-T. Sturm; M. von Renesse. "Transport Inequalities, Gradient Estimates, Entropy and Ricci Curvature". *Commun. Pure Appl. Math.* 58 (2005), 923–940.

[72] M.M. Alexandrino; R.G. Bettiol. *Lie Groups and Geometric Aspects of Isometric Actions*. Springer, 2015.

[73] E. Meckes. *Random Matrix Theory of classical compact groups*. Cambridge University Press, 2018.

[74] J.M. Lee. *Introduction to smooth manifolds*. Springer, 2013.

[75] B. Baumgartner; H. Narnhofer. "Analysis of quantum semigroups with GKS-Lindblad generators: I. Simple generators". *J. Phys. A.* 41 (2008), 065201.

[76] B. Baumgartner; H. Narnhofer. "Analysis of quantum semigroups with GKS-Lindblad generators: II. General". *J. Phys. A.* 41 (2008), 395303.

[77] W. Fulton; J. Harris. *Representation Theory*. Springer, 2004.

[78] V. V. Albert; L. Jiang. "Symmetries and conserved quantities in Lindblad equations". *Phys. Rev. A* 89 (2014), 022118.

[79] A. M. van den Brink; K. Young. "Jordan blocks and generalized bi-orthogonal bases: realizations in open wave systems". *J. Phys. A.* 34 (2001), 2607–2624.

[80] E. Hewitt; K.A. Ross. *Abstract Harmonic Analysis I*. Springer, 1979.

[81] A. Klenke. *Probability Theory - a Comprehensive Course*. Springer, 2020.

[82] W. Rudin. *Real and complex Analysis*. McGraw-Hill, 1987.

[83] C. Villani. *Optimal transport, old and new*. Springer, 2008.

[84] A.B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2009.

[85] H. Holden; B. Øksendal; J. Ubøe; T. Zhang. *Stochastic Partial Differential Equations - A Modeling, White Noise Functional Approach*. Springer, 2010.

[86] J. Bezanson; A. Edelman; S. Karpinski; V.B. Shah. "Julia: A Fresh Approach to Numerical Computing". *SIAM Rev.* 59 (2017), 65–98.

[87] H.-H. Kuo. “White Noise Approach To Stochastic Integration”. *J. Multivar. Anal* 24 (1988), 218–236.

[88] B. Simon. *Representation Theory of finite and compact groups*. Vol. 10. American Mathematical Society, 1991.

[89] Z. Zimborás; R. Zeier; T. Schulte-Herbrüggen; D. Burgarth. “Symmetry Criteria for quantum simulability of effective interactions”. *Phys. Rev. A* 92 (2015), 042309.

[90] E. Wong; M. Zakai. “On the convergence of ordinary Integrals to stochastic Integrals”. *Ann. Math. Stat.* 36 (1965), 1560–1564.

[91] J.E. Humphreys. *Introduction to Lie algebras and their representations*. Springer, 1972.

[92] F. Caravenna; G. Giacomin; M. Gubinelli. “Large scale behaviour of semiflexible heteropolymers”. *Ann. I. H. Poincare-Pr.* 6.1 (2010), 97–118.

[93] J. Bedrossian; A. Blumenthal; S. Punshon-Smith. “A regularity method for lower bounds on the Lyapunov exponent for stochastic differential equations”. *Invent. Math.* (2022), 429–516.

[94] A. Thalmeier. “Geometry of subelliptic diffusions”. *Geometry, Analysis and Dynamics on sub-Riemannian Manifolds*. EMS Series of Lectures in Mathematics. EMS, 2016.

[95] J. Eisert; D. Hangleiter; N. Walk; I. Roth; D. Markham; R. Parekh; U. Chabaud; E. Kashefi. “Quantum certification and benchmarking”. *Nat. Rev. Phys.* 2 (2020), 382–390.

[96] R. Harper; I. Hincks; C. Ferrie; S. Flammia; J. Wallman. “Statistical analysis of randomized benchmarking”. *Phys. Rev. A* 99 (2019), 052350.