arXiv:0802.4195v2 [quant-ph] 9 Dec 2008

Gradient Flows for Optimisation and Quantum Control: Foundations and Applications

T. Schulte-Herbrüggen* and S.J. Glaser
Department of Chemistry, Technical University Munich, Lichtenbergstrasse 4, D-85747 Garching, Germany

G. Dirr† and U. Helmke
Institute of Mathematics, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany

(Dated: 9th December 2008)

For addressing optimisation tasks on finite dimensional quantum systems, we give a comprehensive account on the foundations of gradient flows on Riemannian manifolds including new applications to quantum control: we extend former results on unitary groups to closed subgroups with tensor-product structure, where the finest product partitioning consists of purely local unitary operations. Moreover, the framework is kept sufficiently general for setting up gradient flows on (sub-)manifolds, Lie (sub-)groups, and (reductive) homogeneous spaces. Relevant convergence conditions are discussed, in particular for gradient flows on compact and analytic manifolds. This part of the paper is meant to serve as foundation for some recent and new achievements, and as setting for further research.

Exploiting the differential geometry of quantum dynamics under different scenarios helps to provide highly useful algorithms: (a) On an abstract level, gradient flows may establish the exact upper bounds of pertinent quality functions, i.e. upper bounds reachable within the underlying manifold of the state space dynamics; (b) in a second stage referring to a concrete experimental setting, gradient flows on the space of piecewise constant control amplitudes in $\mathbb{R}^m$ may be set up to yield (approximations to) optimal control for quantum devices under realistic conditions.

Illustrative examples and new applications are given, such as figures of merit on the subgroup of local unitary action on n qubits relating to distance measures of pure-state entanglement. We establish the correspondence to best rank-1 approximations of higher order tensors and show applications from quantum information, where our gradient flows on the subgroup of local unitary operations provide a numerically stable alternative to tensor-SVD techniques.

Keywords: constrained optimisation in quantum state-space manifolds, Riemannian gradient flows and algorithms, double-bracket flows, quantum optimal control, tensor SVD.

AMS Subject Classification 2000: 49-02, 49R50; 53-02, 53Cxx; 65Kxx; 81V70; 90C30; 15A18, 15A69.

PACS Numbers: 02.30.Yy, 02.40.Ky, 02.40.Vh, 02.60.Pn; 03.67.-a, 03.67.Lx, 03.65.Yz, 03.67.Pp

Contents

I. Introduction 2
Outline 3

II. Overview 4
Flows and Dynamical Systems 4
Gradient Flows for Optimisation 4
Discretised Gradient Flows 5
Gradient-Based Methods for Optimal Control 6
Reachability and Controllability 8
Settings of Interest 8

III. Theory: Gradient Flows 9
A. Gradient Flows on Riemannian Manifolds 9
Convergence of Gradient Flows 10
Restriction to Submanifolds 10
Analysing Critical Points by the Hessian 10
Discretised Gradient Flows 11
B. Gradient Flows on Lie Groups 11
Invariant Metrics 12
Gradient Flows with Respect to an Invariant Metric 13
Geodesics with Respect to an Invariant Metric 13
C. Gradient Flows on Homogeneous Spaces 14
Coset Spaces 14
Orbit Theorems and Homogeneous Spaces 15
Reductive Homogeneous Spaces 16
Naturally Reductive Homogeneous Spaces and Geodesics 17
Adjoint Orbits 18
Gradient Flows on Riemannian Homogeneous Spaces 19
D. Examples 19
Worked Example: $SU(N)$ 20
Gradient Flows on $SU(N)$ 20
Gradient Flows on the Local Subgroup $SU_{\mathrm{loc}}(2^n)$ 21
Double-Bracket Flows on Naturally Reductive Homogeneous Spaces 23
Gradient Flows on Naturally Reductive Coset Spaces 26

*Electronic address: tosh@ch.tum.de
†Electronic address: dirr@mathematik.uni-wuerzburg.de

IV. Applications to Quantum Control and Quantum Information 27
A. A Geometric Measure of Pure-State Entanglement 27
B. Generalised Local Subgroup Optimisation 27
Bipartite Systems and Relations to Singular-Value Decompositions 27
Multipartite Systems and Relations to Best Rank-1 Approximations of Higher Order Tensors 28
Numerical Results 29
C. Locally Reversible Interaction Hamiltonians 30
Joint Local Reversibility 30
Pointwise Local Reversibility 31
D. Constrained Optimisation in Quantum Control: Intrinsic vs Penalty Approach 31

V. CONCLUSIONS 34

VI. APPENDIX SECTION 35

Appendix A: Fréchet Differentials and Tangent Maps 35

Appendix B: Polarisation Procedure 35

Acknowledgments 36

References 36

I. INTRODUCTION

Controlling quantum systems offers a great potential for performing computational tasks or for simulating the behaviour of other quantum systems (which are difficult to handle experimentally) or classical systems [1, 2], when the complexity of a problem reduces upon going from a classical to a quantum setting [3]. Important examples are known in quantum computation, quantum search and quantum simulation. Most prominently, there is the exponential speed-up by Shor's quantum algorithm of prime factorisation [4, 5], which on a general level relates to the class of quantum algorithms [6, 7] solving hidden subgroup problems in an efficient way [8]. In Grover's quantum search algorithm [9, 10] one still finds a quadratic acceleration compared to classical approaches [11]. Recently, the simulation of quantum phase-transitions [12] has shifted into focus [13, 14].

Among the generic tools needed for advances in quantum technology, for a survey see, e.g., [15], quantum control plays a major role. Its key concern is not only to find (optimal) control strategies for quantum dynamical systems such that a certain performance index is maximised (typical examples being quantum gate fidelities, efficiencies of state transfer or coherence transfer, as well as distances related to Euclidean entanglement measures) but, moreover, to develop constructive ways for implementing controls under realistic experimental settings. Usually, such figures of merit depend on terminal conditions and running cost, like time or energy, cf. Section II. In quantum control, however, important classes of performance indices are completely determined by some quality function and its value at the system's final state.

Since realistic quantum systems are mostly beyond analytical tractability, numerical methods are often indispensable. A good strategy is to proceed in two steps: (a) firstly, by exploring the possible gains on an abstract and computationally cheap level, i.e. by maximising the quality function either over the entire state space or over the set of possible states—the so-called reachable set; (b) secondly, by going into optimising the experimental controls ('pulse shapes') in a concrete setting. However, (b) is often computationally expensive and highly sensitive, as it actually consists of solving an infinite dimensional constrained variational problem.

By merely depending on the geometry of the underlying state-space manifold the first instance (a) allows for analysing in advance and on an abstract level the limits of what one can achieve in step (b). We therefore refer to (a) as the *abstract optimisation task*. The second instance, in contrast, hinges on introducing the specific time scales of an experimental setting for finding steerings of the quantum dynamical system such that the optima determined in (a) are actually assumed. This is why we term (b) the dynamic *optimal control task*. Certainly, one can approach the entire problem only in terms of (b) and sometimes one is even forced to do so, e.g. if nothing is known about the geometric structure of the reachable set. Yet, the above two-fold strategy may serve to yield a benchmark in (a) for judging the reliability of the numerical results of (b). In both regards, gradient-flow methods will prove to be particularly useful.

In a pioneering paper [16], Brockett introduced the idea of exploiting gradient flows on the orthogonal group for diagonalising real symmetric matrices and for sorting lists. In a series of subsequent papers he extended the concept to intrinsic gradient methods for (constrained) optimisation [17, 18]. Soon these techniques were generalised to Riemannian manifolds, their mathematical and numerical details were worked out [19, 20, 21] and thus they turned out to be applicable to a broad array of optimisation tasks including eigenvalue and singular-value problems, principal component analysis, matrix least-squares matching problems, balanced matrix factorisations, and combinatorial optimisation—for an overview see, e.g., [21, 22]. Implementing gradient flows for optimisation on smooth manifolds, such as unitary orbits, inherently ensures the discretised flow remains within the manifold by virtue of Riemannian exponential map. Alternatively, formulating the optimisation problem on some embedding Euclidean space comes at the expense of additional constraints (e.g. enforcing unitarity) to be taken care of by Lagrange-type or penalty-type techniques. In this sense, gradient flows on manifolds are *intrinsic* optimisation methods [23], whereas *extrinsic* optimisations on the embedding spaces require non-linear projective techniques in order to stay on the (constraint) manifold.

Using the differential geometry of matrix manifolds has thus become a field of active research. For new developments (however without exploiting the Lie structure to

the full extent) see, e.g., [24]. — Even beyond manifolds, gradient flows have recently been described for metric spaces with applications of probability theory [25].

Meanwhile, gradient flows and their discrete numerical integration schemes have also proven powerful tools of optimisation in quantum dynamics. This holds in both of the two types of tasks: (a) for exploring the maxima of pertinent quality functions on the reachable set of a quantum system, e.g. on the unitary group and its orbits [26, 27, 28] and (b) for arriving at concrete experimental steerings (i.e. 'pulse shapes') actually achieving the quality limits established in (a) under given experimental conditions for closed [29, 30, 31] or open [32, 33, 34] systems. Recently we thus gave upper limits for time complexities of implementing unitary modules [30], or for exploring the potential of quantum compilation with complex instruction sets [35].

Moreover, in view of using gradient techniques for unifying variational approaches to ground-state calculations [36], it will be useful to exploit a common framework of gradient flows on Riemannian manifolds as well as projective techniques on their tangent spaces. To this end, we also show how gradient flows can readily be restricted, e.g., from Lie groups $\mathbf{G}$ to any closed subgroup $\mathbf{H}$, in particular any closed subgroup of $SU(N)$. Since quantum dynamics often takes place in a subspace of the entire Hilbert space so that long-range entangling correlations can be neglected on the basis of area laws (see, e.g. [37, 38, 39]), truncating the Hilbert space to the pertinent subspaces is tantamount to representing dynamics of large systems. For instance, unitary networks use consecutive partitionings into different subgroups that can be applied to efficiently compute ground states of large-scale quantum systems with a cost increasing polynomially with system size while retaining sufficiently good approximations. Current approaches of truncating the full-scale Hilbert space into the according subspaces include matrix-product states (MPS) [40, 41] of density-matrix renormalisation groups (DMRG) [42, 43], quantum cellular automata with Margolus partitionings [44], projected entangled pair states (PEPS) [45] weighted graph states (WGS) [46], multi-scale entanglement renormalisation approaches (MERA) [47], string-bond states (SBS) [48] as well as combinations of different techniques [36, 49]. — It is noteworthy, however, that in many-particle physics gradient flows for diagonalising Hamiltonians were re-introduced independently of Brockett's work [16] by Wegner [50] and were further elaborated on again independently of the monography by Helmke and Moore [21] or the one by Bloch [22] in the tract by Kehrein [51]. Suffice this to illustrate the need for making the mathematical methods available to the physics community in a comprehensive way.

Another field of applications of restricting flows to closed subgroups of $SU(N)$ is entanglement of multipartite quantum systems [52, 53]: we present a connection from gradient flows on the subgroup of local unitary operations to best rank–1 approximations of higher order tensors as well as a relation to tensor-SVDs. They are of importance, e.g., in view of optimisation of entanglement witnesses [54]. Gradient flows on partitionings of the full unitary group are anticipated to be of use also for classifying multi-partite systems according to their mutual separability, an example being three-tangles of GHZ-type and W-type states [55, 56, 57].

Moreover, with the framework of treating gradient flows on Riemannian manifolds being very general, we will also show how they can be carried over to homogeneous spaces that do no longer form Lie groups themselves. Standard examples are coset spaces $\mathbf{G}/\mathbf{H}$, i.e. the quotient of a Lie group $\mathbf{G}$ by a closed (yet not necessarily normal) subgroup $\mathbf{H}$. In particular, naturally reductive homogeneous space are in the focus of interest. The well-known double-bracket flows will be demonstrated to form a special case precisely of this kind.

Though gradient flows on the set of control amplitudes can be seen as another instance of flows on Riemannian manifolds, our paper does not primarily focus on optimal control. The goal is rather to give a comprehensive account of the foundations of gradient flows—and thus the justification for some recent developments—as well as to present new flows for intrinsic or extrinsic constraints and new schemes of flows on reductive homogeneous spaces. Terms are kept general enough to trigger future developments, since we elucidate the necessary requirements for implementing gradient-based optimisation methods in different geometric settings: Riemannian manifolds and submanifolds, Lie groups and homogeneous spaces.

A separate paper on open quantum systems [58] sets up a formal approach within the framework of *Lie semigroups* accounting for Markovian quantum evolutions (or Markovian channels). There we also show the current limits of abstract optimisation over reachable sets specifically arising in open systems. The differential geometry of the set of all completely positive, trace-preserving invertible maps is analysed in the framework of *Lie semigroups*. In particular, the set of all Kossakowski-Lindblad generators is retrieved as its tangent cone (Lie wedge). Moreover, it shows how the Lie-semigroup structure corresponds to the Markov properties recently studied in terms of divisibility [59]. It illustrates why *abstract optimisation tasks for open systems* are much more intricate than in the case of closed system, while *dynamic optimal control tasks for open systems* can be handled completely analogously. It specifies algebraic conditions for time-optimal controls to be the method of choice in open systems. Finally it draws perspectives to new algorithmic approaches on semigroup orbits combining (abstract) knowledge of the respective Lie wedges with elements of numerical optimal control.

Outline

To begin with, we consider flows on (Riemannian) manifolds and recall some basic terminology on dynami-

cal systems and Riemannian geometry. Then the aim is to provide the differential geometric tools for setting up gradient flows in different scenarios ranging from optimisation over the entire unitary group to subgroups (e.g. of local actions) or homogeneous spaces. Finally we give a number of applications including worked examples.

More precisely, the paper is organised as follows: Section II draws a general sketch of dynamical systems and flows on manifolds including issues of reachability and controllability. It provides the manifold setting for gradient-flow-based algorithms like steepest ascent, conjugate gradients, Jacobi-type, and Newton methods.

A detailed analysis is then given in Section III, where (1) we resume the general preconditions for gradient flows on smooth manifolds. In particular, we recall the role of a Riemannian metric that allows for identifying the cotangent bundle T^*M with the tangent bundle TM. Large parts of the foundations can be found in [21, 24, 60], but here we additionally provide a comprehensive overview of the interplay between Riemannian geometry, Lie groups, and (reductive) homogeneous spaces. (2) We give examples of gradient flows on compact Lie groups as well as their closed subgroups. (3) In view of further developments, we address gradient flows on several types of reductive homogeneous spaces: Cartan-like, naturally reductive ones and merely reductive ones. In particular, double-bracket flows turn out as gradient flows on naturally reductive homogeneous spaces. (4) Examples interdispersed in the main text illustrate the relevance in a plethora of different settings.

Section IV is dedicated to specific applications in quantum control and quantum information. (1) We show how gradient flows on the subgroup of local unitaries $SU_{\mathrm{loc}}(2^n)$ in n qubits do not only provide a valuable tool in witness optimisation, but relate to generalised singular-value decompositions (SVD), namely the tensor-SVD. Here, our gradient flows yield an alternative to common algorithms for best rank-1 approximations of higher-order tensors, e.g. higher-order power methods (HOPM) or higher-order orthogonal iteration (HOOI). (2) Flows on $SU_{\mathrm{loc}}(2^n)$ also serve as a convenient tool to decide whether Hamiltonian interactions can be time-reversed solely by local unitary manipulations thus complementing the algebraic assessment given in [28]. (3) Optimisation tasks with (i) intrinsic and (ii) extrinsic constraints are addressed by tailored gradient flows on the respective subgroups (i) or with Lagrange-type penalties (ii). By including practical applications and worked examples we illustrate the ample range of problems to which gradient flows on manifolds provide valuable solutions. — To this end, we start out by an extended overview on techniques on (Riemannian) manifolds with particular emphasis on gradient techniques.

II. OVERVIEW

Flows and Dynamical Systems

In this paper, we treat various optimisation tasks for quantum dynamical systems in a common framework, namely by gradient flows on smooth manifolds. Let M denote a smooth manifold, e.g. the unitary orbit of all quantum states relating to an initial state X_0. By a continuous-time dynamical system or a *flow* one defines a smooth map

$$\Phi : \mathbb{R} \times M \to M \tag{1}$$

such that for all states $X \in M$ and times $t, \tau \in \mathbb{R}$ one has

$$\begin{aligned} \Phi(0, X) &= X \\ \Phi(\tau, \Phi(t, X)) &= \Phi(t + \tau, X) \quad . \end{aligned} \tag{2}$$

Since these equations hold for any $X \in M$, one gets the operator identity

$$\Phi_\tau \circ \Phi_t = \Phi_{t+\tau} \tag{3}$$

for all $t, \tau \in \mathbb{R}$, thus showing the flow acts as a one-parameter *group*, and for positive times $t, \tau \geq 0$ as a one-parameter *semigroup of diffeomorphisms* on M.

Gradient Flows for Optimisation

Now, the general idea for optimising a scalar quality function on a smooth manifold M (which might either arise naturally or from including smooth equality constraints, *vide infra*) by dynamical systems is as follows: Let $f : M \to \mathbb{R}$ be a smooth quality function on M. The differential of $f : M \to \mathbb{R}$ is a mapping $Df : M \to T^*M$ of the manifold to its cotangent bundle T^*M, while the gradient vector field is a mapping $\operatorname{grad} f : M \to TM$ to its tangent bundle TM. So the *gradient* of f at $X \in M$, denoted $\operatorname{grad} f(X)$, is the vector in $T_X M$ uniquely determined by

$$\mathrm{D}\, f(X) \cdot \xi = \langle \operatorname{grad} f(X) | \xi \rangle_X \quad \text{for all } \xi \in T_X M. \tag{4}$$

Here, the scalar product $\langle \cdot | \cdot \rangle_X$ plays a central role: it allows for identifying $T_X^* M$ with $T_X M$. The pair $\big(M, \langle \cdot | \cdot \rangle\big)$ is called a *Riemannian manifold* with *Riemannian metric* $\langle \cdot | \cdot \rangle$. In view of gradient flows, the convenience of Riemannian manifolds lies in the fact that by duality in particular the differential $\mathrm{D}\, f(X)$ of f at X can be identified with a tangent vector of $T_X M$.

Then, the flow $\Phi : \mathbb{R} \times M \to M$ determined by the ordinary differential equation

$$\dot{X} = \operatorname{grad} f(X) \tag{5}$$

is termed *gradient flow*. Formally, it is obtained by integrating Eqn. (5), i.e.

$$\Phi(t, X) = \Phi(t, \Phi(0, X)) = X(t) \,, \tag{6}$$

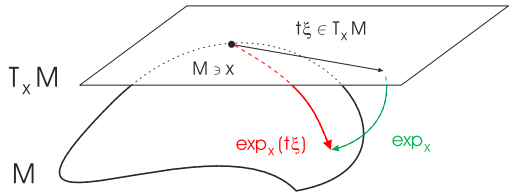

Figure 1: (Colour online) The Riemannian exponential $\exp_X$ is a smooth map taking the tangent vector $t\xi \in T_X M$ at $X \in M$ to $\exp_X(t\xi) \in M$. By varying $t \in \mathbb{R}$, it yields the unique geodesic with initial value $X \in M$ and 'initial velocity' $\xi \in T_X M$.

where $X(t)$ denotes the unique solution of Eqn. (5) with initial value $\mathrm{X}(0) = \mathrm{X}$. Observe this ensures that f does increase along trajectories Φ of the system by virtue of following the gradient direction of f.

Discretised Gradient Flows

Gradient flows may be envisaged as natural continuous versions of the steepest ascent method for optimising a real-valued function $f : \mathbb{R}^m \to \mathbb{R}$ by moving along its gradient $\operatorname{grad} f \in \mathbb{R}^m$, i.e.

Steepest Ascent Method

$$X_{k+1} = X_k + \alpha_k \operatorname{grad} f(X_k), \tag{7}$$

where $\alpha_k \geq 0$ is an appropriate step size.

Here, the right hand side of Eqn. (7) does make sense, as the manifold $M = \mathbb{R}^m$ coincides with its tangent space $T_X M = \mathbb{R}^m$ containing $\operatorname{grad} f(X)$. Clearly, a generalisation is required as soon as M and $T_X M$ are no longer identifiable. This gap is filled by the *Riemannian exponential map* defined by

$$\exp_X : T_X M \to M\,, \quad \xi \mapsto \exp_X(\xi) \tag{8}$$

so that $t \mapsto \exp_X(t\xi)$ describes the unique geodesic with initial value $X \in M$ and 'initial velocity' $\xi \in T_X M$ as illustrated in Fig. 1.

If the manifold M carries the structure of a matrix Lie group $\mathbf{G}$, we may identify the tangent space element $\xi \in T_X \mathbf{G}$ with ΩX, where Ω is itself an element of the Lie algebra $\mathfrak{g}$, i.e. the tangent space at the unity element $\mathfrak{g} = T_{\mathbb{1}}\mathbf{G}$. Moreover, if the Lie-group structure matches with the Riemannian framework in the sense that the metric is bi-invariant (as will be explained in more detail later), then the Riemannian exponential of $\xi = \Omega X$ can readily be calculated explicitly. This is done in three steps by (i) right translation with the inverse group element X^{-1}, (ii) taking the conventional exponential map of the Lie algebra element Ω, (iii) right translation with the group element X as summarised in the following diagram

$$\begin{array}{ccc} \xi = \Omega X \in T_X \mathbf{G} & \xrightarrow{\exp_X} & e^{\Omega} X \in \mathbf{G} \\ {\scriptstyle R_{X^{-1}}}\downarrow & & \uparrow{\scriptstyle R_X} \\ \Omega \in \mathfrak{g} & \xrightarrow{\exp} & e^{\Omega} \in \mathbf{G}\,. \end{array} \tag{9}$$

Next, the gradient system (5) will be integrated (to sufficient approximation) by a discrete scheme that can be seen as an *intrinsic Euler step method*. This can be performed by way of the Riemannian exponential map, which is to say straight line segments used in the standard method are replaced by geodesics on M. This leads to the following integration scheme which is well-defined on any Riemannian manifold M.

1. *Riemannian Gradient Method*

$$X_{k+1} := \exp_{X_k}\big(\alpha_k \operatorname{grad} f(X_k)\big) \tag{10}$$

where $\alpha_k \geq 0$ denotes a step size appropriately selected to guarantee convergence, cf. Section III.

For matrix Lie groups $\mathbf{G}$ with bi-invariant metric, Eqn. (10) simplifies to

1'. *Gradient Method on a Lie Group*

$$X_{k+1} := \exp\big(\alpha_k \operatorname{grad} f(X_k)\; X_k^{-1}\big) X_k\,, \tag{11}$$

where $\exp : \mathfrak{g} \to \mathbf{G}$ denotes the conventional exponential map.

In either case, the iterative procedure can be pictured as follows: at each point $X_k \in M$ one evaluates $\operatorname{grad} f(X_k)$ in the tangent space $T_{X_k} M$. Then one moves via the Riemannian exponential map in direction $\operatorname{grad} f(X_k)$ to the next point X_{k+1} on the manifold so that the quality function f improves, $f(X_{k+1}) \geq f(X_k)$, as shown in Fig. 2.

The steepest ascent approach just outlined is most basic for addressing abstract optimisation tasks intrinsically. Other intrinsic iterative schemes exploiting the underlying Riemannian geometry like conjugate gradients, Jacobi-type methods or Newton's method can be obtained similarly. For an introduction to these more advanced topics beyond the subsequent sketch see, e.g., Refs. [19, 61, 62].

2. *Conjugate Gradient Method*

$$\begin{aligned} X_k^{l+1} &:= \operatorname*{argmax}_{t \geq 0} f\big(\exp_{X_k^l}(t\,\Omega_k^l)\big) \\ X_{k+1}^0 &:= X_k^n, \end{aligned} \tag{12}$$

$$\Omega_k^l := \begin{cases} \operatorname{grad} f(X_k^l) & \text{for } l = 0 \\ \operatorname{grad} f(X_k^l) + \alpha_k^l \Pi_{X_{k^{l-1}}, X_k^l}(\Omega_{k-1}^l) & \\ & \text{for } l = 1, \ldots, n-1, \end{cases}$$

where α_k^l is a real parameter and $\Pi_{X,Y}(\Omega)$ denotes the parallel transport of Ω along the geodesic from X to Y.

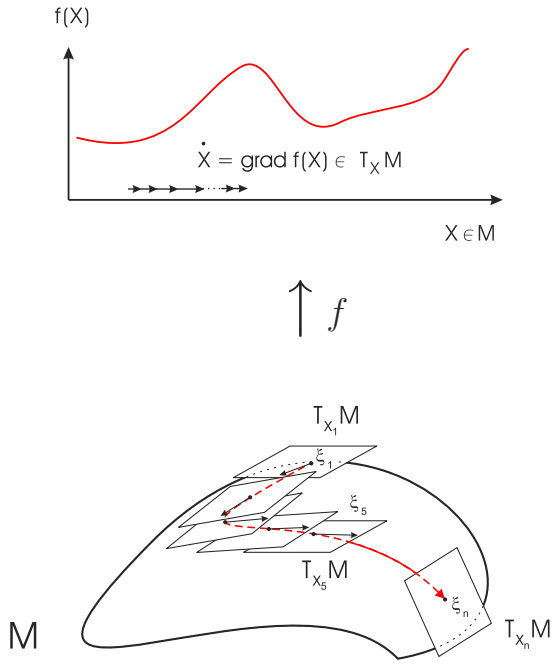

Figure 2: (Colour online) Abstract optimisation task: the quality function $f : M \to \mathbb{R}, X \mapsto f(X)$ (top trace) is driven into a (local) maximum by following the gradient flow $\dot{X} = \operatorname{grad} f(X)$ on the manifold M (lower trace).

3. *Jacobi-Type Method*

$$\begin{aligned} X_k^{l+1} &:= \operatorname*{argmax}_{t\in\mathbb{R}} f\Big(\exp_{X_k^l}\big(t\,\Omega_l(X_k^l)\big)\Big) \\ X_{k+1}^0 &:= X_k^s, \end{aligned} \tag{13}$$

where $\Omega_0, \Omega_1, \ldots, \Omega_{s-1}$ are vector fields such that $\Omega_0(X), \Omega_1(X), \ldots, \Omega_{s-1}(X)$ span $T_X M$ for all $X \in M$. The integer s is called *sweep length*.

4. *Newton's Method*

$$X_{k+1} := \exp_{X_k}\Big(-\big(\operatorname{Hess} f(X_k)\big)^{-1} \operatorname{grad} f(X_k)\Big), \tag{14}$$

where $\operatorname{Hess} f(X)$ denotes the Hessian of f at X.

Gradient-Based Methods for Optimal Control

Up to now we have addressed optimisation tasks over state spaces forming abstract manifolds M—hence the term *abstract optimisation task* (AOT). In this subsection, we briefly describe how gradient methods like the one in Ref. [29] arise in the context of *optimal control tasks* (OCT). These algorithms are of practical relevance, as in many applications, the entire state space M of a physical system (Σ) evolving under some internal dynamics and external controls is not accessible.

Now the general optimal control task amounts to finding the time course of controls to achieve a maximum of a given quality functional l, which may depend on the time evolution of the controls and the state—so-called 'running costs'—as well as on the terminal state. Here, we consider only quality functionals l which are determined by the terminal state of (Σ), i.e. l is given by some smooth quality function f on the state space M. Note, however, that this is in general not a restriction as running costs can often be reduced to an end point condition, see below.

Thus, the optimal control task reduces to finding controls that drive the initial state X_0 to a maximum of

$$f|_{\mathrm{Reach}(X_0)}, \tag{15}$$

i.e. to a maximum of f restricted to what is known as the *reachable set* $\mathrm{Reach}(X_0)$ of X_0. It is the set of all possible states the system can be driven into within positive time. Moreover, the set of all states which can be reached just in time T will be denoted by $\mathrm{Reach}(X_0, T)$. Thus

$$\mathrm{Reach}(X_0) := \bigcup_{T\geq 0} \mathrm{Reach}(X_0, T). \tag{16}$$

A more detailed discussion on this topic will be given in the next subsection.

Choosing an (equidistant) sufficiently small partitioning $0 = t_0, \ldots, t_{n_p} = T$ of the interval $[0, T]$ and assumimg that the set of all states which can be reached in time $T \geq 0$ by *piecewise constant* controls is dense in $\mathrm{Reach}(X_0, T)$, the optimal control task translates into a maximisation over the set

$$M_{cp} := \mathrm{Step}(t_0, \ldots, t_{n_p}; \mathbb{R}^{n_c}) \tag{17}$$

of all piecewise constant controls on $[0, T]$ with *fixed* switching points $t_1, \ldots, t_{n_p-1}$. It is important to note that M_{cp} is a vector space isomorphic to $\mathbb{R}^{n_c \cdot n_p}$, where n_c and n_p are the respective numbers of controls and subintervals.

In order to express the effect of a piecewise constant control on the final state of (Σ), we define a mapping

$$\phi_T : M_{cp} \to M, \quad Y \mapsto X(T, Y), \tag{18}$$

where $X(t, Y) \equiv X(t, Y_1, \ldots, Y_{n_p})$ denotes the unique solution of the control system (Σ) with initial value $X(0, Y) = X_0$ and piecewise constant control $Y_j \in \mathbb{R}^{n_c}$ on $[t_{j-1}, t_j]$. Hence, the impact of the controls on the quality function is described by the composition

$$g_T := f \circ \phi_T : M_{cp} \to \mathbb{R} \quad . \tag{19}$$

Often, the map $\phi_T : M_{cp} \to M$ is 'almost' *surjective* in the sense that the interior of its image $\phi_T(M_{cp})$ relative to $\mathrm{Reach}(X_0, T)$ is open and dense. Therefore, ϕ_T can be interpreted as a highly non-regular over–parametrisation of the set $\mathrm{Reach}(X_0, T)$. Now, any standard non-linear optimisation tool, e.g. an Euclidian gradient method allows to compute the (local) maximum of g_T and thus to

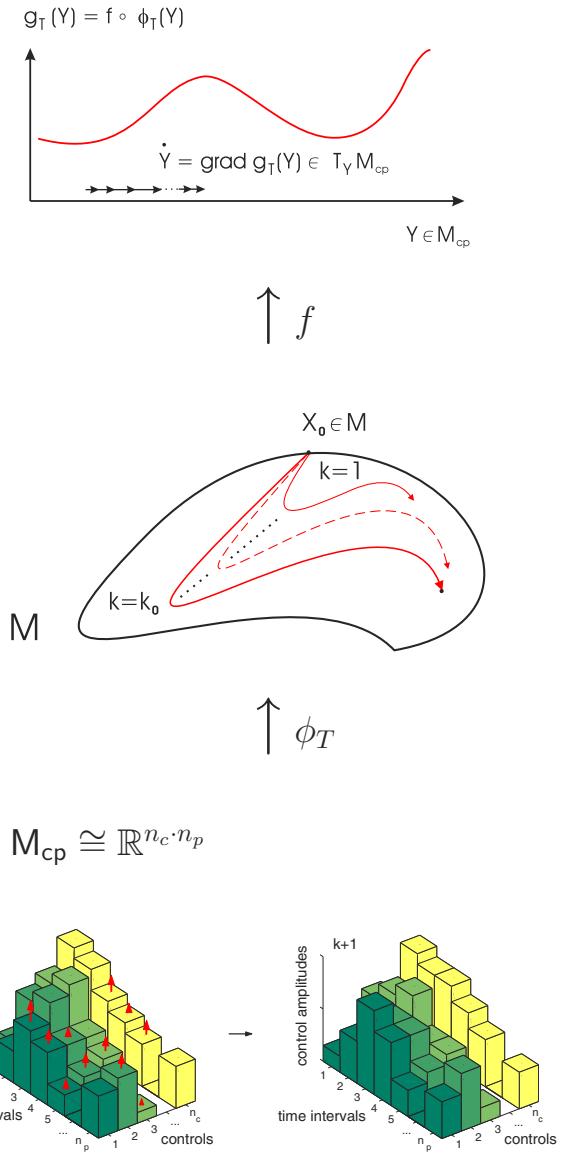

Figure 3: (Colour online) Optimal control task: the quality function $f : M \to \mathbb{R}, X \mapsto f(X)$ is driven into a (local) maximum on the reachable set $\mathrm{Reach}(X_0) \subseteq M$ by following an implicit procedure (intermediate panel). It is brought about by a gradient flow $\dot{Y} = \mathrm{grad}\, g_T(Y)$ on the level of experimental control amplitudes $Y \in M_{cp}$ (lower traces) so that a standard steepest ascent method applies.

numerically evaluate the overall quality function g defined by

$$g(T) := \max_{T' \leq T} \Big(\max_{Y} g_{T'}(Y) \Big),$$

where Y varies over all piecewise constant controls in $\mathrm{Step}(t_0, \ldots, t_{n'_p}, \mathbb{R}^{n_c})$ with $t_0 = 0$ and $t_{n'_p} = T'$. Hence, for solving the OCT, at last one has to increment $T \geq 0$ until $g(T)$ remains constant. Moreover, *time-optimal* solutions of the OCT are obtained while searching for the minimal time $T^* > 0$ such that $g(T^*)$ is equal to the maximum of g.

In order to be more explicit, assume that the system (Σ) is given by a time-continuous control system, i.e. by an (ordinary) differential equation

$$(\Sigma) \qquad \dot{X} = F(X, u), \quad u \in \mathbb{R}^m \tag{20}$$

depending on some (unrestricted) control parameters $u \in \mathbb{R}^m$. Thus the general OCT amounts to maximising a functional

$$l(X, u, T) := f(X(T)) + \int_0^T h\big(X(t), u(t)\big)\mathrm{d}t \tag{21}$$

under the constraint Eqn. (20). Here, the integral term in Eqn. (21) represents the so-called 'running costs'. We therefore suppose that h vanishes. This, however, is not really a restriction as mentioned above. By augmenting an auxiliary state variable and state equation to (Σ) one can reduce 'running costs' to a terminal state condition, cf. [63, 64]. Moreover, note that in the setting of time-continuous control systems the final scanning for the optimal value of g can be accomplished by introducing another auxiliary optimisation parameter $u_0 > 0$ in (Σ) such that

$$(\Sigma_0) \qquad \dot{X} = u_0 F(X, u), \quad (u_0, u) \in \mathbb{R}^{m+1} \tag{22}$$

while keeping the time interval $[0, T]$ fixed, e.g. $T = 1$. Further numerical methods from optimal control can be found in e.g. [63, 65, 66, 67, 68, 69]. Some of these more involved techniques require *Pontryagin's maximum principle*, which can be viewed as Lagrange-multiplier method for constrained variational problems. These techniques often come at the cost of additional boundary-value problems.

To sum up, the quality function $f|_{\mathrm{Reach}(X_0)}$ is driven into a (local) maximum by an implicit procedure in the sense that it is not explicitly defined on the reachable set $\mathrm{Reach}(X_0) \subseteq M$. Rather it is the result of a gradient flow

$$\dot{Y} = \mathrm{grad}\, g_T(Y). \tag{23}$$

on the level of control amplitudes $Y \in M_{cp} \cong \mathbb{R}^{n_p \cdot n_c}$ so that one finally gets the discretised version

$$Y_{k+1} = Y_k + \alpha_k \,\mathrm{grad}\, g_T(Y). \tag{24}$$

Thus the iterative scheme (24) reads as a standard steepest ascent method. These ideas are illustrated in Fig. 3; they are exploited in numerical optimal-control methods like the GRAPE algorithm [29].

Finally, note that the OCT relates to an AOT if the reachable set $\mathrm{Reach}(X_0)$ itself has the geometric structure of a submanifold N of the state-space manifold M. However, in general this is not the case and would have to be ensured *a priori*. Therefore, if one seeks to reduce the OCT to an AOT, geometric control and reachability aspects come into play as discussed in the next

subsection. For instance, in the Hamiltonian unitary evolution of a finite dimensional closed quantum system, the closure of the reachable set takes always the form of a *group orbit* of some initial state [70], e.g. $\mathcal{O}(\rho_0) := \{U\rho_0 U^\dagger \mid U \text{ unitary}\}$. Its Riemannian geometry is well understood. In open dissipative systems, however, the dynamics is governed, e.g., by a Markovian quantum Master equation and thus by a *semigroup* of completely positive operators. As will be illustrated in Ref. [58], it is much harder to give an explicit characterisation of their reachable sets. Apart from utterly simple scenarios, an abstract approach is often unviable in dissipative systems, and thus implicit methods by optimal control techniques may become indispensible.

Reachability and Controllability

In the following, some general remarks on reachable sets and controllability will clarify the previous, slightly sloppy introduction of these term. For simplicity, let (Σ) denote a smooth *control system* on the manifold M, i.e. a familiy of (ordinary) differential equations

$$(\Sigma) \qquad \dot{X} = F(X,u), \quad u \in \mathcal{U} \subset \mathbb{R}^m \tag{25}$$

with control parameters $u \in \mathcal{U}$ and smooth vector fields $F_u := F(\cdot, u)$ on M. While the vector fields F_u are assumed to be time-independent, the controls are allowed to vary in time. For convenience, the resulting control function $t \mapsto u(t) \in \mathcal{U}$ is denoted again by u. Moreover, the set $\mathcal{U}$ of all *admissible controls functions* is supposed to contain at least all piecewise constant ones.

For $u \in \mathcal{U}$, we refer to $X(t, X_0, u)$ as the unique solution of (25) with initial value X_0. Therewith, the *reachable set* of X_0 is defined by

$$\mathrm{Reach}(X_0) := \bigcup_{0 \le T} \mathrm{Reach}(X_0, T). \tag{26}$$

Here $\mathrm{Reach}(X_0, T)$ denotes the set of all states which can be reached in time T, i.e.

$$\mathrm{Reach}(X_0, T) := \big\{X(T, X_0, u) \in M \mid u \in \mathcal{U}\big\}, \tag{27}$$

cf. (16). The system (Σ) is said to be *controllable*, if $\mathrm{Reach}(X_0) = M$ for all $X_0 \in M$, i.e. if for each pair (X_0, Y_0) there exists an admissible control u and a time $T \ge 0$ such that $X(T, X_0, u) = Y_0$.

In general, it is hard to decide whether a given system (Σ) is controllable or not. However, for dynamics expressed on some Lie group $\mathbf{G}$, the situation is much easier. Let $(\Sigma_{\mathbf{G}})$ be a *bilinear* or, equivalently, a *right invariant, control-affine* system on a matrix Lie group $\mathbf{G}$ with Lie algebra $\mathfrak{g}$, i.e.

$$(\Sigma_{\mathbf{G}}) \qquad \dot{X} = (A_0 + \sum_{j=1}^{m} u_j A_j)X, \quad u \in \mathbb{R}^m \tag{28}$$

with drift $A_0 \in \mathfrak{g}$ and control directions $A_j \in \mathfrak{g}$. For *compact* Lie groups $\mathbf{G}$, a simple algebraic test for controllability is known: If the *system Lie algebra*

$$\mathfrak{s} := \langle A_0, \ldots, A_m \rangle_{\mathrm{Lie}} \tag{29}$$

generated by $A_0, \ldots, A_m$ via nested commutators coincides with $\mathfrak{g}$, then the corresponding system $(\Sigma_{\mathbf{G}})$ is controllable, cf. [71, 72]. In particular, there exists a finite time $T' > 0$, such that the entire group $\mathbf{G}$ can be reached from any initial point $X_0 \in \mathbf{G}$ within this time, i.e.

$$\mathbf{G} = \bigcup_{0 \le T \le T'} \mathrm{Reach}(X_0, T) =: \mathrm{Reach}(X_0, \le T') \tag{30}$$

for all $X_0 \in \mathbf{G}$.

In contrast, for *non-compact* groups $\mathbf{G}$, which are indispensible for the description of dissipation in open quantum systems, the situation gets more involved. Here, $\mathfrak{s} = \mathfrak{g}$ implies only *accessibility* of $(\Sigma_{\mathbf{G}})$, i.e. that all reachable sets $\mathrm{Reach}(X_0)$ have non-empty interior. This follows from a more general result on smooth non-linear control systems—the so-called *Lie algebra rank condition* (LARC)

$$\big\{F(X) \mid F \in \langle F_u \mid u \in \mathcal{U}\rangle_{\mathrm{Lie}}\big\} = T_X M, \tag{31}$$

where $\langle F_u \mid u \in \mathcal{U}\rangle_{\mathrm{Lie}}$ denotes the Lie subalgebra of vectors fields generated by F_u, $u \in \mathcal{U}$ via Lie bracket operation, cf. [64, 73]. Note that for right invariant vector fields on $\mathbf{G}$, the Lie bracket essentially coincides with the commutator such that (31) boils down to $\mathfrak{s} = \mathfrak{g}$. Moreover, by exploiting the identity

$$\mathrm{Reach}(\mathbb{1}, T_1) \cdot \mathrm{Reach}(\mathbb{1}, T_2) = \mathrm{Reach}(\mathbb{1}, T_1 + T_2), \tag{32}$$

one can show that $\mathrm{Reach}(\mathbb{1})$ is always a *Lie subsemigroup* of $\mathbf{G}$ [58]. Here, a *subsemigroup* is a subset $\mathbf{S} \subset \mathbf{G}$ which contains the unity and is closed under multiplication, i.e. $\mathbb{1} \in \mathbf{S}$ and $\mathbf{S} \cdot \mathbf{S} \subseteq \mathbf{S}$. However, the geometry of subsemigroups is rather subtle compared to Lie subgroups and therefore at present not amenable to intrinsic optimisation methods, as will be shown in more detail in a separate paper dwelling on open systems [58].

Settings of Interest

In terms of reachability, there are different scenarios that structure the subsequent line of thought: we start out with *fully controllable* or *operator controllable* quantum systems [27, 71, 72, 74, 75, 76, 77] represented as spin- or pseudo-spin systems. Then, neglecting decoherence, to any initial state represented by its density operator A, the entire unitary orbit $\mathcal{O}(A) := \{UAU^{-1} \,|\, U \text{ unitary}\}$ can be reached [77]. In systems of n qubits (*e.g.* spin-$\frac{1}{2}$ particles), this is the case under the following mild conditions [26, 27, 78]: (1) the qubits have to be inequivalent, i.e., distinguishable and selectively addressable, and (2) they have to be pairwise

coupled (e.g., by Ising or Heisenberg-type interactions), where the coupling topology may take the form of any connected graph. In other instances not the entire (unitary) group, but just a subgroup $\mathbf{K}$ can be reached. This is the case, if the coupling topology is not a connected graph or the n qubits cannot be addressed by separate controls.

Otherwise, the system itself can be fully controllable, but the the focus of interest may be reduced: e.g., the subgroup $\mathbf{K} = SU_{\mathrm{loc}}(2^n) := SU(2) \otimes SU(2) \otimes \cdots \otimes SU(2)$ of (possibly fast) local actions on each qubit is of interest to study local reachability, or whether an effective multi-qubit interaction Hamiltonian is locally reversible in the sense of Hahn's spin echo [28]. Or, one may ask what is the Euclidean distance of some pure state to the nearest point on the local unitary orbit of a pure product state. This may be useful when optimising entanglement witnesses [54]. Likewise, one may address other than the finest partitioning of the entire unitary group, e.g. $\mathbf{K} = SU(2^{n_1}) \otimes \cdots \otimes SU(2^{n_j}) \otimes \cdots \otimes SU(2^{n_r}) \subset SU(2^n)$, where $\sum_{j=1}^{r} n_j = n$.

Another type of reduction arises not by restriction to a subgroup $\mathbf{H}$, but by the fact that the quality function of interest f is *equivariant*, i.e. constant on cosets $\mathbf{H}G$. Consider, for instance, a fully controllable system where f is equivariant with respect to the closed subgroup $\mathbf{H} \subset \mathbf{G}$. Then, it may be favourable to transfer the optimisation problem from the original Lie group $\mathbf{G}$ to the homogeneous space $\mathbf{G}/\mathbf{H}$.

III. THEORY: GRADIENT FLOWS

Gradient systems are a standard tool of Riemannian optimisation for maximizing smooth quality functions on a manifold M. Thus the manifold structure of M arises either naturally by the problem itself or by smooth equality-constraints imposed on a previously unconstrained problem. Note that in general inequality-constraints would entail manifolds with a boundary—and thus are a much more subtle issue not to be developed any further here.

The case $M = \mathbb{R}^m$—sometimes refered to as the unconstrained case—is well-known and can be found in many texts on ordinary differential equations or nonlinear programming, cf. [79, 80, 81, 82]. However, gradient systems on abstract Riemannian manifolds provide a rather new approach to constrained optimisation problems. Although the resulting numerical algorithms are in general only linearly convergent, their global behaviour is often much better then the global behaviour of locally quadratic methods.

Textbooks combining the different areas of Riemannian geometry, gradient systems and constrained optimisation are quite rare. The best choices to our knowledge are [21, 60]. For further reading we also suggest the papers [18, 19, 61]. Nevertheless, most of the material which is necessary to understand the intrinsic optimisation approach applied in Section IV is scattered in many different references. For the reader's convenience, we therefore review the basic ideas on these topics. First, we discuss the general setting on Riemannian manifolds, then we proceed with Lie groups and finally summarize some more advanced results on homogeneous spaces. For standard definitions and terminology from Riemannian geometry we refer to any modern text on this subject such as [83, 84, 85].

A. Gradient Flows on Riemannian Manifolds

In the following, let M denote a finite dimensional smooth manifold M with tangent and cotangent bundles $\mathrm{T}M$ and T^*M, respectively. Moreover, let M be equipped with a *Riemannian metric* $\langle\cdot|\cdot\rangle$, i.e. with a scalar product $\langle\cdot|\cdot\rangle_X$ on each tangent space $\mathrm{T}_X M$ varying smoothly with $X \in M$. More precisely, $\langle\cdot|\cdot\rangle$ has to be a smooth, positive definite section in the bundle of all symmetric bilinear forms over M. Such sections always exist for finite dimensional smooth manifolds, cf. [86]. The pair $\big(M, \langle\cdot|\cdot\rangle\big)$ is called a *Riemannian manifold*.

Let $f : M \to \mathbb{R}$ be a smooth quality function on M with differential $Df : M \to T^*M$. Then the *gradient* of f at $X \in M$, denoted by $\operatorname{grad} f(X)$, is the vector in $\mathrm{T}_X M$ uniquely determined by the equation

$$\mathrm{D}\, f(X) \cdot \xi = \langle \operatorname{grad} f(X) | \xi \rangle_X \tag{33}$$

for all $\xi \in T_X M$. Equation (33) naturally defines a vector field on M via

$$\operatorname{grad} f \;:\; M \to \mathrm{T}M, \quad X \mapsto \operatorname{grad} f(X) \tag{34}$$

called the *gradient vector field* of f. The corresponding ordinary differential equation

$$\dot{X} = \operatorname{grad} f(X), \tag{35}$$

and its flow are referred to as the *gradient system* and the *gradient flow* of f, respectively.

Obviously, the *critical points* of $f : M \to \mathbb{R}$ coincide with the equilibria of the gradient flow. Moreover, the quality function f is monotonically increasing along solutions $X(t)$ of (35), i.e. the real-valued function $t \mapsto f(X(t))$ is monotonically increasing in t, as

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} f\big(X(t)\big) &= \langle \operatorname{grad} f(X(t)) | \dot{X}(t) \rangle_{X(t)} \\ &= || \operatorname{grad} f(X(t)) ||^2_{X(t)} \geq 0. \end{aligned}$$

Here $||\cdot||_X$ denotes the norm on $\mathrm{T}_X M$ induced by $\langle\cdot|\cdot\rangle_X$, i.e. $||\xi||_X := \sqrt{\langle\xi|\xi\rangle_X}$ for all $\xi \in T_X M$.

Convergence of Gradient Flows

Recall that the asymptotic behaviour (for $t \to +\infty$) of a solution of (35) is characterised by its *ω-limit set*

$$\omega(X_0) := \bigcap_{0<t<t^+(X_0)} \overline{\{X(\tau, X_0) \mid t \leq \tau < t^+(X_0)\}},$$

where $\overline{\{\cdots\}}$ denotes the closure of the set $\{\cdots\}$ and $X(t, X_0)$ the unique solution of (35) with initial value $X(0) = X_0$ and positive escape time $t^+(X_0) > 0$.

The following result gives a sufficient condition for solutions of Eqn. (35) to converge to the set of critical points of f.

Proposition III.1. *If f has compact a superlevel set, i.e. if the sets $\{X \in M \mid f(X) \geq C\}$ are compact for all $C \in \mathbb{R}$, then any solution of Eqn. (35) exists for $t \geq 0$ and its ω-limit set is a non-empty compact and connected subset of the set of critical points of f.*

Proof. Since solutions of Eqn. (35) are monotonically increasing in t, the compact sets $\{X \in M \mid f(X) \geq C\}$ are positively invariant, i.e. invariant for $t \geq 0$ under the gradient flow of Eqn. (35). Thus, the assertion follows from standard results on ω-limit sets and Lyapunov theory, cf. [21, 80]. □

Although, Proposition III.1 guarantees that $\omega(X_0)$ is contained in the set of critical points of f, this does not imply convergence to a critical point. Indeed, there are smooth gradient systems which exhibit solutions converging only to the set of critical points, cf. [87]. The next two results provide sufficient conditions for convergence to a single critical point under different settings. In particular, Threorem III.1 yields a powerful tool for analysing real analytic gradient systems.

Corollary III.1. *If f has compact superlevel sets and if all critical points are isolated, then any solution of (35) converges to a critical point of f for $t \to +\infty$.*

Proof. This is an immediate consequence of Proposition III.1. □

Theorem III.1 (Łojasiewicz)**.** *If $\big(M, \langle\cdot|\cdot\rangle\big)$ and f are real analytic, then all non-empty ω-limit sets $\omega(X_0)$ of Eqn. (35) are singletons, i.e. $\omega(X_0) \neq \emptyset$ implies that $X(t, X_0)$ converges to a single critical point X^* of f for $t \to +\infty$.*

Proof. The main argument is based on Łojasiewicz's inequality which says that in a neighbourhood of X^* an estimate of the type

$$|f(X)|^p \leq C \| \operatorname{grad} f(X)\|$$

for some $p < 1$ and $C > 0$ holds. A complete proof can be found in [88, 89]. □

Restriction to Submanifolds

Now, consider the restriction of f to a smooth submanifold $N \subset M$. Obviously, the Riemannian metric $\langle\cdot|\cdot\rangle$ on M restricts to a Riemannian structure on N. Thus $\big(N, \langle\cdot|\cdot\rangle|_{\mathrm{T}N}\big)$ constitutes a Riemannian manifold in a canonical way. Moreover, the equality $\mathrm{D}\, f|_N(X) = \mathrm{D}\, f(X)|_{\mathrm{T}_X N}$ immediately implies that the gradient of the restriction $f|_N$ at $X \in N$ is given by the orthogonal projection of $\operatorname{grad} f(X)$ onto $\mathrm{T}_X N$, i.e.

$$\operatorname{grad} f|_N(X) = P_X\big(\operatorname{grad} f(X)\big), \tag{36}$$

where P_X denotes the orthogonal projector onto $\mathrm{T}_X N$. Hence the gradient system of $f|_N$ on an arbitrary submanifold N is well-defined and reads

$$\dot{X} = P_X\big(\operatorname{grad} f(X)\big). \tag{37}$$

Analysing Critical Points by the Hessian

Subsequently, we address the problem, how to define and compute the *Hessian* of f, as its knowledge is essential for a deeper insight of (35). For instance, the stability of critical points is determined by its eigenvalues or the computation of explicit discretisation schemes, preserving the convergence behaviour of (35), can be based on it, cf. [16, 21].

At critical points $X^* \in M$ of f, the *Hessian* is given by the symmetric bilinear form $\operatorname{Hess} f(X^*) : \mathrm{T}_{X^*}M \times \mathrm{T}_{X^*}M \to \mathbb{R}$,

$$\begin{aligned}\operatorname{Hess} f(X^*)(\xi, \eta) \;:=\; & \\ (\mathrm{D}\,\varphi(X^*)\xi)^\top \operatorname{Hess}(f \circ \varphi^{-1})\big(\varphi(X^*)\big)\, \mathrm{D}\,\varphi(X^*)\eta,&\end{aligned} \tag{38}$$

where φ is any chart around X^* and $\operatorname{Hess}(f \circ \varphi^{-1})$ denotes the ordinary Hesse matrix of $f \circ \varphi^{-1}$. It is straightforward to show that (38) is independent of φ. Equivalently, $\operatorname{Hess} f(X^*)$ is uniquely determined by

$$\operatorname{Hess} f(X^*)(\xi, \xi) := \frac{\mathrm{d}^2(f \circ \alpha)}{\mathrm{d}t^2}\Big|_{t=0}, \tag{39}$$

where α is any smooth curve with $X^* = \alpha(0)$ and $\dot{\alpha}(0) = \xi$. While the remaining values of $\operatorname{Hess} f(X^*)$ can be obtained by a standard polarisation argument, i.e. via the formula

$$\begin{aligned}2\operatorname{Hess} f(X^*)(\xi, \eta) = \operatorname{Hess} f(X^*)(\xi + \eta, \xi + \eta)& \\ -\operatorname{Hess} f(X^*)(\xi, \xi) - \operatorname{Hess} f(X^*)(\eta, \eta)&\end{aligned} \tag{40}$$

However, the previous definition does not apply to regular points of f. In general, one has to establish the concept of *geodesics*, cf. Remark III.1 More precisely, the Hessian of f at an arbitrary point $x \in M$ is defined by

$$\operatorname{Hess} f(X)(\xi, \xi) := \frac{\mathrm{d}^2(f \circ \gamma)}{\mathrm{d}t^2}\Big|_{t=0}, \tag{41}$$

where γ is the unique geodesic with $X = \gamma(0)$ and $\dot\gamma(0) = \xi$, cf. [137]. Again, the remaining values can be computed by (40). As usual, we associate to $\mathrm{Hess} f(X)$ a unique selfadjoint linear operator $\mathbf{Hess} f(X) : \mathrm{T}_X M \to \mathrm{T}_X M$ such that

$$\langle \xi | \mathbf{Hess} f(X) \eta \rangle_X = \mathrm{Hess} f(X)(\xi, \eta) \tag{42}$$

holds for all $\xi, \eta \in \mathrm{T}_X M$. It is called the *Hessian operator* of f at $X \in M$.

Remark III.1. In modern textbooks, geodesics are defined via *linear connections* on M, cf. [85, 90]. For Riemannian manifolds M, however, it is possible to introduce (Riemannian) geodesics as curves of minimal arc length. Both concepts coincide (locally) if we choose the so-called *Riemannian* or *Levi-Civita connection* on M. Unfortunately, the computation of geodesics is in general a highly non-trivial problem. However, on compact Lie groups their calculation is much easier as we will see at the end of Section III B.

The above concepts yield the following generalisation of a familiar result from elementary calculus.

Theorem III.2. *Let M be a Riemannian manifold and let X_* be a critical point of the quality function $f : M \to \mathbb{R}$. If $\mathrm{Hess} f(X^*)$ or, equivalently, $\mathbf{Hess} f(X)$ are negative definite, then X^* is a strict local maximum of f.*

Proof. Use local coordinates, then the result follows straightforwardly from Eqn. (38). □

In general, (asymptotic) stability of an equilibrium $X^* \in M$ of (35) may dependent on the Riemannian metric $\langle \cdot | \cdot \rangle$. However, the property of being a strict local maximum or an isolated critical point of a smooth function f is obviously not up to the choice of any Riemannian metric. Therefore, the following result shows that in fact certain (asymptotically) stable equilibria $X^* \in M$ of (35) are independent of the Riemannian metric.

Theorem III.3. *(a) If $X^* \in M$ is a strict local maximum of f, then X^* is a stable equilibrium of (35). In particular, for any neighbourhood U of X^* there exists a neighbourhood V of X^* such that the ω-limit sets $\omega(x_0)$ are non-empty and contained in U for all $x_0 \in V$.*

(b) If $X^* \in M$ is a strict local maximum and an isolated critical point of f, then X^* is an asymptotically stable equilibrium of (35). In particular, there is a neighbourhood V of X^* such that $\omega(x_0) = \{X^*\}$ for all $X_0 \in V$, i.e. all solutions $X(t, X_0)$ with initial value $X_0 \in V$ converge to X^* for $t \to +\infty$.

Proof. Both assertions follow immediately from classical stability theory by taking f as Lyapunov function, cf. [21, 80] □

Note that the convergence analysis near arbitrary equilibria, i.e. near arbitrary critical points of f is quite subtle and may depend on the Riemannian metric, cf. [91].

Discretised Gradient Flows

Finally, we approach the problem of finding discretisations of (35) which lead to convergent *gradient ascent methods*. The ideas presented below can be traced back to R. Brockett, cf. [16]. Let

$$\exp_X : \mathrm{T}_X M \to M \tag{43}$$

be the *Riemannian exponential map* at $X \in M$, i.e. $t \mapsto \exp_X(t\xi)$ denotes the unique geodesic with initial value $X \in M$ and initial velocity $\xi \in \mathrm{T}_X M$. Moreover, we assume that M is (geodesically) *complete*, i.e. any geodesic is defined for all $t \in \mathbb{R}$. Hence, (43) is well-defined for the entire tangent bundle $\mathrm{T}M$.

The simplest discretisation approach—a scheme that can be seen as an *intrinsic Euler step method*—leads to

Riemannian Gradient Method

$$X_{k+1} := \exp_{X_k}\big(\alpha_k \operatorname{grad} f(X_k)\big) \tag{44}$$

where α_k denotes an appropriate step size.

In order to guarantee convergence of (44) to the set of critical points, it is sufficient to apply the *Armijo rule* [82]. An alternative to Armijo's rule provides the step size selection suggested by Brockett in [18], see also [21]. Convergence to a single critical point is a more subtle issue. If $\big(M, \langle \cdot | \cdot \rangle\big)$ and f are analytic, and the step sizes are chosen according to a version of the first *Wolfe-Powell* condition for Riemannian manifolds, then pointwise convergence holds. A detailed proof can be found in [92].

B. Gradient Flows on Lie Groups

In the following, we apply the previous results to Lie groups and Lie subgroups. However, to fully exploit Lie-theoretic tools, the Riemannian structure and the group structure have to match, i.e. the metric $\langle \cdot | \cdot \rangle$ has to be invariant under the group action. For basic concepts and results on Lie Groups and their Riemannian geometry we refer to [84, 93]. In particular, we recommend the book of Arvanitoyeorgos [94] for a rather condensed overview.

Let $\mathbf{G}$ denote a finite dimensional *Lie group*, i.e. a group which carries a smooth manifold structure such that the group operations are smooth mappings. For notational convenience we will assume that $\mathbf{G}$ can be represented as a (closed) *matrix Lie group*, i.e. as an (embedded) Lie subgroup of some *general linear group* $GL(N, \mathbb{K})$ of invertible $N \times N$-matrices over $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$.

Remark III.2. According to a well-known result by Cartan, a subgroup $\mathbf{G} \subset GL(N, \mathbb{K})$ is an (embedded) Lie subgroup, i.e. a smooth submanifold of $GL(N, \mathbb{K})$, if and only if it is closed in $GL(N, \mathbb{K})$, cf. [95]. Note, however, that there is a subtle difference between embedded and immersed Lie subgroups. Moreover, not every abstract

Lie group admits a faithful representation as a matrix Lie group. Nevertheless, the class of matrix Lie groups is rich enough for all of our subsequent applications. For more details on these topics we refer to [96].

Invariant Metrics

A Lie group $\mathbf{G}$ can be endowed in a canonical way with a Riemannian metric $\langle\cdot|\cdot\rangle$. Let $\mathfrak{g} := \mathrm{T}_{\mathbb{1}}\mathbf{G}$ be the Lie algebra of $\mathbf{G}$, i.e. the tangent space of $\mathbf{G}$ at the unity $\mathbb{1}$. From the fact that the right multiplication $r_H : \mathbf{G} \to \mathbf{G}$ and left multiplication $l_H : \mathbf{G} \to \mathbf{G}$ are diffeomorphisms of $\mathbf{G}$ for all $H \in \mathbf{G}$, it follows

$$\mathrm{T}_H\mathbf{G} = \mathfrak{g}H = H\mathfrak{g} \tag{45}$$

for all $H \in \mathbf{G}$. Now, let $(\cdot|\cdot)$ be any scalar product on $\mathfrak{g}$. Then

$$\langle g|h\rangle_G := (gG^{-1}|hG^{-1}) \tag{46}$$

for all $G \in \mathbf{G}$ and $g, h \in \mathrm{T}_G\mathbf{G}$ yields a *right invariant* metric on $\mathbf{G}$, where right invariance stands for

$$\langle g|h\rangle_G = \langle gH|hH\rangle_{GH} \tag{47}$$

for all $G, H \in \mathbf{G}$ and $g, h \in \mathrm{T}_G\mathbf{G}$. Thus, right multiplication r_H represents an *isometry* of G. In the same way, one could obtain *left invariant* metrics on $\mathbf{G}$.

Remark III.3. In an abstract setting, one has to replace (45) by

$$\mathrm{T}_H\mathbf{G} = \mathrm{D}\, r_H(\mathbb{1})\mathfrak{g} = \mathrm{D}\, l_H(\mathbb{1})\mathfrak{g} \tag{48}$$

for all $H \in \mathbf{G}$, where $\mathrm{D}\, r_H$ and $\mathrm{D}\, l_H$ denote the tangent maps of r_H and l_H, respectively. For a matrix Lie group, however, the respective tangent maps are given by $\mathrm{D}\, r_H(G)\xi = \xi H$ and $\mathrm{D}\, l_H(G)\xi = H\xi$ for all $G \in \mathbf{G}$ and $\xi \in \mathrm{T}_G\mathbf{G}$. Hence (48) reduces to (45).

The construction of *bi-invariant*, i.e. right and left invariant metrics is much more subtle and in general even impossible. To summarise the basic results on this topic we need some further terminology. The *adjoint* maps $\mathrm{Ad} : \mathbf{G} \to GL(\mathfrak{g})$ and $\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$ are defined by

$$\mathrm{Ad}_G\, h := GhG^{-1}$$

and

$$\mathrm{ad}_g\, h := [g, h] := gh - hg$$

for all $G \in \mathbf{G}$ and all $g, h \in \mathfrak{g}$, where $GL(\mathfrak{g})$ and $\mathfrak{gl}(\mathfrak{g})$ denote the set of all automorphisms and, respectively, endomorphisms of $\mathfrak{g}$. Note both notations $\mathrm{ad}_g\, h$ and $[g, h]$ are used interchangeably in the literature. A bilinear form $(\cdot|\cdot)$ on $\mathfrak{g}$ is called

(a) $\mathrm{Ad}_{\mathbf{G}}$-*invariant* if the identity

$$(g|h) = (\mathrm{Ad}_G\, g|\, \mathrm{Ad}_G\, h) \tag{49}$$

is satisfied for all $g, h \in \mathfrak{g}$ and $G \in \mathbf{G}$.

(b) $\mathrm{ad}_{\mathfrak{g}}$-*invariant* if the identity

$$(\mathrm{ad}_g\, h|k) = -(h|\, \mathrm{ad}_g\, k) \tag{50}$$

is satisfied for all $g, h, k \in \mathfrak{g}$.

Proposition III.2. *The following statements are equivalent:*

(a) There exists a bi-invariant Riemannian metric $\langle\cdot|\cdot\rangle$ on $\mathbf{G}$.

(b) There exists an $\mathrm{Ad}_{\mathbf{G}}$-invariant scalar product $(\cdot|\cdot)$ on $\mathfrak{g}$.

Moreover, each of the statements (a) and (b) imply

(c) There exists an $\mathrm{ad}_{\mathfrak{g}}$-invariant scalar product $(\cdot|\cdot)$ on $\mathfrak{g}$.

If $\mathbf{G}$ is also connected, then (c) is equivalent to (a) and (b), respectively,

Proof. The equivalence (a) $\Longleftrightarrow$ (b) follows easily by exploiting Eqn. (47) at $G = \mathbb{1}$. Moreover, applying (b) to a one-parameter subgroup $t \mapsto \exp(tg)$ and taking the derivative with respect to $t = 0$ yields (c). The implication (c) $\Longrightarrow$ (b) is obtained in the same way, i.e. by taking the derivative of

$$t \mapsto \big(\mathrm{Ad}_{\mathrm{e}^{tg}}\, h|\, \mathrm{Ad}_{\mathrm{e}^{tg}}\, k\big).$$

with respect to t, cf. [94]. Note, however, that this implies $\mathrm{Ad}_{\mathbf{G}}$-invariance only on the connected component of the unity. Therefore, connectedness is necessary for the implication (c) $\Longrightarrow$ (b) as counter-examples show. □

Now, the main result on the existence of bi-invariant metrics reads as follows.

Theorem III.4. *A connected Lie group $\mathbf{G}$ admits a bi-invariant Riemannian metric if and only if $\mathbf{G}$ is the direct product of a compact Lie group $\mathbf{G}_0$ and an abelian one, which is isomorphic to some $(\mathbb{R}^m, +)$, i.e. $\mathbf{G} \cong \mathbf{G}_0 \times \mathbb{R}^m$*

Proof. Cf. [84, 97]. □

Finally, we focus on a special class of Lie groups. A connected Lie group $\mathbf{G}$ is called *semisimple* if the Killing form, i.e. the bilinear form $(g, h) \mapsto \kappa(g, h) := \mathrm{tr}(\mathrm{ad}_g\, \mathrm{ad}_h)$ is non-degenerate on $\mathfrak{g}$. Most prominent representatives of this class are $SL(N, \mathbb{R})$, $SL(N, \mathbb{C})$, $SO(N, \mathbb{R})$ and $SU(N)$. More on semisimple Lie groups and their algebras can be found in [93, 95].

Theorem III.5. *(a) If $\mathbf{G}$ is semisimple then the Killing form κ defines an $\mathrm{ad}_{\mathfrak{g}}$-invariant bilinear form on $\mathfrak{g}$.*

(b) If $\mathbf{G}$ is semisimple and compact then $-\kappa$ defines an $\mathrm{ad}_{\mathfrak{g}}$-invariant scalar product on $\mathfrak{g}$. Thus $-\kappa$ induces a bi-invariant Riemannian metric on $\mathbf{G}$.

Proof. Cf. [94, 95]. □

Next, we study gradient flows on $\mathbf{G}$ or on a closed subgroup $\mathbf{H} \subset \mathbf{G}$ with respect to an invariant metric $\langle\cdot|\cdot\rangle$. Therefore, let $f : \mathbf{G} \to \mathbb{R}$ be a smooth quality function and let $\varphi : \mathbf{G} \to \mathbf{G}$ be any diffeomorphism. Using the identity

$$\operatorname{grad}(f \circ \varphi)(G) = \big(\mathrm{D}\,\varphi(G)\big)^* \operatorname{grad} f\big(\varphi(G)\big) \tag{51}$$

for all $G \in \mathbf{G}$, where $(\cdot)^*$ denotes the adjoint operator, we obtain by the right invariance of the metric

$$\operatorname{grad} f(G) = \operatorname{grad}(f \circ r_G)(\mathbb{1})G \tag{52}$$

for all $G \in \mathbf{G}$. Hence

$$\dot{G} = \operatorname{grad} f(G). \tag{53}$$

can be rewritten as

$$\dot{G} = \operatorname{grad}(f \circ r_G)(\mathbb{1})G. \tag{54}$$

Thus, the gradient flow of f is completely determined by the mapping $G \mapsto \operatorname{grad}(f \circ r_G)(\mathbb{1}) \in \mathfrak{g}$. To study its asymptotic behaviour of Eqn. (53) we can apply the results of the previous section. For instance, for compact Lie grops we have.

Corollary III.2. *Let $\mathbf{G}$ be a compact Lie group with a right invariant Riemannian metric $\langle\cdot|\cdot\rangle$ and let $f : \mathbf{G} \to \mathbb{R}$ be a real analytic quality function. Then any solution of Eqn. (53) converges to a critical point of f for $t \to +\infty$.*

Proof. This follows immediately from Proposition III.1 and Theorem III.1, as the pair $\big(\mathbf{G}, \langle\cdot|\cdot\rangle\big)$ constitutes a real analytic Riemannian manifold whenever the metric $\langle\cdot|\cdot\rangle$ is invariant, cf. [98]. □

Now, let $\mathbf{H}$ be a closed subgroup of $\mathbf{G}$. By Remark III.2, we know that $\mathbf{H}$ is actually an (embedded) submanifold of $\mathbf{G}$. Therefore, the gradient flow of $f|_{\mathbf{H}}$ with respect to $\langle\cdot|\cdot\rangle|_H$ is well-defined and can be given explicitly via the orthogonal projectors P_H, cf. (37). However, for an invariant metric the computation of P_H simplifies considerably, as all calculations can be carried out on the Lie algebra $\mathfrak{g}$ of $\mathbf{G}$.

Lemma III.1. *Let $\mathbf{G}$ be a Lie group with a right invariant Riemannian metric $\langle\cdot|\cdot\rangle$ and let $\mathbf{H}$ be a closed subgroup of $\mathbf{G}$. Furthermore, let $\mathfrak{g}$ and $\mathfrak{h}$ their corresponding Lie algebras and denote by $P_{\mathfrak{h}}$ the orthogonal projection of $\mathfrak{g}$ onto $\mathfrak{h}$. Then the orthogonal projection P_H in (36) is given by*

$$P_H(gH) := P_{\mathfrak{h}}(g)H \tag{55}$$

for all $gH \in \mathrm{T}_H\mathbf{G}$.

Proof. This is a straightforward consequence of the identity $\mathrm{T}_H\mathbf{H} = \mathfrak{h}H$ and the right invariance of $\langle\cdot|\cdot\rangle$. □

According to (37), (52) and (55), the gradient flow of $f|_H$ finally reads

$$\dot{H} = P_{\mathfrak{h}}\big(\operatorname{grad}(f \circ r_H)(\mathbb{1})\big)H. \tag{56}$$

The remainder of this subsection is devoted to the issue: How to compute geodesics and the Hessian of a smooth quality function with respect to an invariant metric. The main results for the forthcoming applications are summarised in Theorem III.6(b) and Proposition III.3. For readers with basic differential geometric background we provide some details of the proof which however can be skipped, so as not to lose the thread. First, we need some further notation. Let

$$\mathcal{X}^r_g : G \mapsto gG \quad \text{and} \quad \mathcal{X}^l_g : G \mapsto Gg \tag{57}$$

be the *right* and *left invariant vector fields* on $\mathbf{G}$ which are uniquely determined by $\mathcal{X}^r_g(\mathbb{1}) = g$ and $\mathcal{X}^l_g(\mathbb{1}) = g$, respectively. Moreover, let $L_{\mathcal{X}}(\cdot)$ denote the Lie derivative with respect to the vector field $\mathcal{X}$, i.e. for a smooth function $f : \mathbf{G} \to \mathbb{R}$ one has

$$L_{\mathcal{X}}(f)(G) := \mathrm{D}\,f(G) \cdot \mathcal{X}(G).$$

On vector fields $\mathcal{Y}$, the action of $L_{\mathcal{X}}(\cdot)$ is given by

$$L_{\mathcal{X}}(\mathcal{Y})(G) := - \lim_{t\to\infty} \frac{\big(\mathrm{D}\,\Phi_{\mathcal{X}}(t,G)\big)^{-1} \cdot \mathcal{Y}\big(\Phi_{\mathcal{X}}(t,G)\big) - \mathcal{Y}(G)}{t},$$

where $\Phi_{\mathcal{X}}(t,\cdot)$ denotes the corresponding flow of $\mathcal{X}$.

Next, we recall two basic facts from differential geometry which play a key role for the proof of Theorem III.6. The first one shows that the set of right/left invariant vector fields is invariant under Lie derivation, cf. [93, 94]. The second one relates a Riemannian metric of a manifold M with a particular linear connection on M. For more details see e.g. [84].

Fact 1. The Lie derivative of a right/left invariant vector field is again right/left invariant and satisfies

$$L_{\mathcal{X}^r_g}\mathcal{X}^r_h = -\mathcal{X}^r_{[g,h]} \quad \text{and} \quad L_{\mathcal{X}^l_g}\mathcal{X}^l_h = \mathcal{X}^l_{[g,h]}. \tag{58}$$

Fact 2. On any Riemannian manifold M there exists a unique *Riemannian connection* ∇ determined by the properties

$$L_{\mathcal{X}}\mathcal{Y} = \nabla_{\mathcal{X}}\mathcal{Y} - \nabla_{\mathcal{Y}}\mathcal{X} \tag{59}$$

and

$$\nabla_{\mathcal{X}}\langle\mathcal{Y}|\mathcal{Z}\rangle = \langle\nabla_{\mathcal{X}}\mathcal{Y}|\mathcal{Z}\rangle + \langle\mathcal{Y}|\nabla_{\mathcal{X}}\mathcal{Z}\rangle. \tag{60}$$

Now, combining both facts yields the main result about geodesics on Lie groups.

Theorem III.6. *Let $\mathbf{G}$ be a Lie group with a bi-invariant metric $\langle\cdot|\cdot\rangle$ and let ∇ denote the unique Riemannian connection on $\mathbf{G}$ induced by $\langle\cdot|\cdot\rangle$.*

(a) For right/left invariant vector fields the Riemannian connection ∇ is given by

$$\nabla_{\mathcal{X}^r_g}\mathcal{X}^r_h = -\frac{1}{2}\mathcal{X}^r_{[g,h]} \quad \textit{and} \quad \nabla_{\mathcal{X}^l_g}\mathcal{X}^l_h = \frac{1}{2}\mathcal{X}^l_{[g,h]}. \tag{61}$$

(b) The geodesics through any $G \in \mathbf{G}$ are of the form $t \mapsto G\exp(tg)$ or $t \mapsto \exp(tg)G$ with $g \in \mathfrak{g}$. In particular, the geodesics through the unity $\mathbb{1}$ are precisely the one-parameter subgroups of $\mathbf{G}$.

Proof. (a) Applying Koszul's identity, cf. [84, 94],

$$\begin{aligned} 2\langle \nabla_{\mathcal{X}}\mathcal{Y}|\mathcal{Z}\rangle &= L_{\mathcal{X}}\langle \mathcal{Y}|\mathcal{Z}\rangle + L_{\mathcal{Y}}\langle \mathcal{Z}|\mathcal{X}\rangle - L_{\mathcal{Z}}\langle \mathcal{X}|\mathcal{Y}\rangle \\ &\quad -\langle \mathcal{X}|L_{\mathcal{Y}}\mathcal{Z}\rangle + \langle \mathcal{Y}|L_{\mathcal{Z}}\mathcal{X}\rangle + \langle \mathcal{Z}|L_{\mathcal{X}}\mathcal{Y}\rangle, \end{aligned}$$

to $\mathcal{X}^r_g, \mathcal{X}^r_h$ and $\mathcal{X}^r_k$ we obtain

$$\begin{aligned} &2\langle \nabla_{\mathcal{X}^r_g}\mathcal{X}^r_h|\mathcal{X}^r_k\rangle = \\ &\quad +\langle \mathcal{X}^r_g|\mathcal{X}^r_{[h,k]}\rangle - \langle \mathcal{X}^r_h|\mathcal{X}^r_{[k,g]}\rangle - \langle \mathcal{X}^r_k|\mathcal{X}^r_{[g,h]}\rangle. \end{aligned}$$

Now Proposition III.2 and Fact 1 implies

$$2\langle \nabla_{\mathcal{X}^r_g}\mathcal{X}^r_h|\mathcal{X}^r_k\rangle = -\langle \mathcal{X}^r_k|\mathcal{X}^r_{[g,h]}\rangle$$

and hence

$$2\nabla_{\mathcal{X}^r_g}\mathcal{X}^r_h = -\frac{1}{2}\mathcal{X}^r_{[g,h]}.$$

Obviously, for left invariant vector fields the same arguments apply.

(b) Let $\gamma(t) := \exp(tg)G$. Part (a) implies that the covariant derivative $\nabla_{\dot\gamma(t)}\dot\gamma(t) = \nabla_{\mathcal{X}^r_g}\mathcal{X}^r_g(\gamma(t))$ of γ vanishes and thus γ represents the unique geodesics through G with 'initial velocity' $\xi = gG$. The same holds for $\widetilde\gamma(t) := G\exp(tg)$, cf. [94] or [93]. □

Observe that the bi-invariance of the metric and the invariance of the vector fields are essential for the above result. For example Eqn. (61) fails, if the Riemannian metric is just right invariant. More details on this topic can be found in [93, 99]

Finally, by Theorem III.6, the Hessian of the restriction $f|_{\mathbf{H}}$ can easily be obtained by restricting the Hessian of f to T$\mathbf{H}$. More precisely, we have.

Proposition III.3. *Let $f : \mathbf{G} \to \mathbb{R}$ be a smooth quality function on a Lie group with bi-invariant metric $\langle\cdot|\cdot\rangle$ and let $\mathbf{H}$ be a closed subgroup. Then the Hessian of $f|_{\mathbf{H}}$ at H is given by*

$$\mathrm{Hess} f|_{\mathbf{H}}(H) = \mathrm{Hess} f(H)\Big|_{\mathrm{T}_H\mathbf{H}\times\mathrm{T}_H\mathbf{H}} \tag{62}$$

Note that in general Eqn. (62) is sheer nonsense unless $\mathbf{H}$ is a Lie subgroup. Counter-examples can be obtained easily for $\mathbf{G} = \mathbb{R}^m$.

C. Gradient Flows on Homogeneous Spaces

The subsequent section on homogeneous spaces is motivated by the following observation, cf. Subsection III D. As before, let $f : \mathbf{G} \to \mathbb{R}$ be a smooth quality function. In many applications f can be decomposed into a function F defined on a smooth manifold M and a (right) *group action* $\alpha : (X, G) \mapsto X \cdot G$ on M such that

$$f(G) := F(X \cdot G) \tag{63}$$

for some fixed $X \in M$. Then we can think of f as defined on the orbit of X. More precisely, let $\widehat f = F|_{\mathcal{O}(X)}$, where $\mathcal{O}(X) := \{X \cdot G \mid G \in \mathbf{G}\}$ denotes the *orbit* of X. Thus,

$$\widehat f(Y) = f(G) \tag{64}$$

for $Y = X \cdot G$ with $G \in \mathbf{G}$. Such quality functions f are called *induced* by F, cf. Subsection III D. By construction, we have

$$\max_{G\in\mathbf{G}} f(G) = \max_{Y\in\mathcal{O}(X)} \widehat f(Y). \tag{65}$$

Moreover, let $\mathbf{H}_X := \{G \in \mathbf{G} \mid X \cdot G = X\}$ denote the *stabiliser* or, equivalently, *isotropy subgroup* of X. Then $\widehat f$ can also be viewed as a function on the *right coset space*

$$\mathbf{G}/\mathbf{H}_X := \{\mathbf{H}_X G \mid G \in \mathbf{G}\}, \tag{66}$$

cf. [138], which is equivalent to say that f is *equivariant* with respect to $\mathbf{H}_X$, i.e.

$$f(G) = f(HG) \tag{67}$$

for all $H \in \mathbf{H}_X$. Therefore, coset space show up quite naturally in optimising equivariant quality functions. Note that passing from $\mathbf{G}$ to $\mathbf{G}/\mathbf{H}_x$ can be rather useful in order to avoid certain degeneracies such as continua of critical points.

Coset Spaces

We first collect the fundamental facts on the differential structure of $\mathbf{G}/\mathbf{H}$, where $\mathbf{H}$ is any closed subgroup of $\mathbf{G}$. Detailed expositions can be found in [93, 94, 100, 101, 102].

Theorem III.7. *Let $\mathbf{G}$ be a Lie group with Lie algebra $\mathfrak{g}$ and let $\mathbf{H} \subset \mathbf{G}$ be a closed subgroup with Lie algebra $\mathfrak{h}$. Moreover, let $\mathfrak{p}$ be any complementary subspace to $\mathfrak{h}$, i.e. $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{p}$. Then the following holds:*

(a) The quotient topology turns the set of right cosets $\mathbf{G}/\mathbf{H} := \{[G] := \mathbf{H}G \mid G \in \mathbf{G}\}$ into a locally compact Hausdorff space.

(b) There exists a unique manifold structure on $\mathbf{G}/\mathbf{H}$ such that the canonical projection $\Pi : \mathbf{G} \to \mathbf{G}/\mathbf{H}$, $G \mapsto [G]$ is a submersion. In particular, the tangent space of $\mathbf{G}/\mathbf{H}$ at $[\mathbb{1}]$ is isomorphic to $\mathfrak{p}$ via the canonical identification $p \mapsto \frac{\mathrm{d}}{\mathrm{d}t}[\exp tp]\big|_{t=0}$ and thus $\dim\mathbf{G}/\mathbf{H} = \dim\mathbf{G} - \dim\mathbf{H}$.

The following statements refer to the unique manifold structure on $\mathbf{G}/\mathbf{H}$ given in part (b).

(c) The Lie group $\mathbf{G}$ *acts smoothly from the right on* $\mathbf{G}/\mathbf{H}$ *via*

$$\big([G'],G\big) \mapsto [G'G] \tag{68}$$

such that

$$\widehat{r}_G : \mathbf{G}/\mathbf{H} \to \mathbf{G}/\mathbf{H}, \quad [G'] \mapsto [G'G] \tag{69}$$

are diffeomorphisms for all $G \in \mathbf{G}$*. Moreover,*

$$\Pi \circ l_G : \mathbf{G} \to \mathbf{G}/\mathbf{H}, \quad G' \mapsto [GG'] \tag{70}$$

are submersions for all $G' \in \mathbf{G}$*. Thus, the tangent space* $\mathrm{T}_{[G]}\mathbf{G}/\mathbf{H}$ *is given by*

$$\begin{aligned} \mathrm{T}_{[G]}\mathbf{G}/\mathbf{H} &= \mathrm{D}\,\widehat{r}_G([\mathbb{1}])\,\mathrm{T}_{[\mathbb{1}]}\mathbf{G}/\mathbf{H} \\ &= \mathrm{D}(\Pi \circ l_G)(\mathbb{1})\mathfrak{g} \\ &= \mathrm{D}(\Pi \circ l_G)(\mathbb{1})(\mathrm{Ad}_{G^{-1}}\,\mathfrak{p}). \end{aligned} \tag{71}$$

(d) Moreover, if $\mathbf{H}$ *is a normal subgroup, i.e.* $G\mathbf{H}G^{-1} = \mathbf{H}$ *for all* $G \in \mathbf{G}$*, then the multiplication* $[G]\cdot[G'] := [GG']$ *is well-defined and yields a Lie group structure on* $\mathbf{G}/\mathbf{H}$*.*

Proof. Cf. [93, 98, 102]. □

The Lie group $\mathbf{G}/\mathbf{H}$ given by Theorem III.7 (d) is called the *quotient Lie group* of $\mathbf{G}$ by $\mathbf{H}$. Moreover, the result provides the possibility to extend the well-known *First Isomorphism Law* to the category of Lie groups.

Theorem III.8. *Let* $\Phi : \mathbf{G} \to \mathbf{G}'$ *be a smooth surjective Lie group homomorphism. Then there exists a well-defined Lie group isomorphism* $\widehat{\Phi} : \mathbf{G}/\mathbf{H} \to \mathbf{G}'$ *with* $\mathbf{H} := \ker\Phi$ *such that the diagram*

$$\begin{array}{ccc} \mathbf{G} & \xrightarrow{\Phi} & \mathbf{G}' \\ {\scriptstyle\Pi}\downarrow & \nearrow_{\widehat{\Phi}} & \\ \mathbf{G}/\mathbf{H} & & \end{array} \tag{72}$$

commutes. Moreover, let $\mathfrak{g}$*,* $\mathfrak{g}'$ *and* $\mathfrak{h}$ *denote the corresponding Lie algebras and let* $\mathfrak{p}$ *be any complementary space to* $\mathfrak{h}$*. Then* $\mathrm{D}\,\Phi(\mathbb{1})$ *is a surjective Lie algebra homomorphism with* $\ker\mathrm{D}\,\Phi(\mathbb{1}) = \mathfrak{h}$ *and commutative diagram*

$$\begin{array}{ccc} \mathfrak{g} & \xrightarrow{\mathrm{D}\,\Phi(\mathbb{1})} & \mathfrak{g}' \\ {\scriptstyle\mathrm{D}\,\Pi(\mathbb{1})}\downarrow & \nearrow_{\mathrm{D}\,\widehat{\Phi}(\mathbb{1})} & \\ \mathfrak{p} \cong \mathfrak{g}/\mathfrak{h}\,. & & \end{array} \tag{73}$$

Proof. Note that $\mathbf{H} = \ker\Phi$ is a closed normal subgroup of $\mathbf{G}$. Thus, by the First Isomorphism Law $\widehat{\Phi}([G]) := \Phi(G)$ for $[G] \in \mathbf{G}/\mathbf{H}$ is a well-defined group isomorphism. Moreover, $\widehat{\Phi}$ is smooth, since Π is a smooth submersion by Theorem III.7. The assertion that $\mathrm{D}\,\Phi(\mathbb{1})$ is a surjective Lie algebra homomorphism, follows easily from the properties of the exponential map. Finally, a straightforward application of the chain rule yields Eqn. (73). □

Next, we analyse the relation between group actions and coset spaces. A smooth right *Lie group action* is a smooth map $\alpha : M \times \mathbf{G} \to M$, $(X,G) \mapsto X\cdot G$ which satisfies

$$(X\cdot G)\cdot H = X\cdot(GH) \quad \text{and} \quad X\cdot\mathbb{1} = X$$

for all $X \in M$ and $G, H \in \mathbf{G}$. The *orbit* of $X \in M$ under the group action α is defined by $\mathcal{O}(X) := \{X\cdot G \mid G \in \mathbf{G}\}$. The action is called *transitive* if $M = \mathcal{O}(X)$ for some and hence for all $X \in M$. Equivalently, one can say that for all $X, Y \in M$ there exists an element $G \in \mathbf{G}$ with $Y = X\cdot G$. Moreover, for $X \in M$ let $\mathbf{H}_X := \{G \in \mathbf{G} \mid X\cdot G = X\}$ denote the *stabiliser* of X and $\alpha_X : \mathbf{G} \to M$ the map $G \mapsto X\cdot G$. Then the *canonical map* $\widehat{\alpha}_X : \mathbf{G}/\mathbf{H}_X \to M$ is defined by $[G] \mapsto X\cdot G$.

Theorem III.9 (Orbit Theorem)**.** *Let* $\mathbf{G}$ *be a Lie group with Lie algebra* $\mathfrak{g}$ *and let* $\alpha : M \times \mathbf{G} \to M$ *be a smooth right action of* $\mathbf{G}$ *on a smooth manifold* M*. Moreover, let* X *be any point in* M*. Then the following statements are satisfied:*

(a) The stabiliser subgroup $\mathbf{H}_X$ *is a closed subgroup of* $\mathbf{G}$*.*

(b) Let $\mathfrak{h}_X$ *be the Lie algebra of* $\mathbf{H}_X$*. Then*

$$\ker\mathrm{D}\,\alpha_X(\mathbb{1}) = \mathfrak{h}_X. \tag{74}$$

In particular, the canonical map $\widehat{\alpha}_X : \mathbf{G}/\mathbf{H}_X \to M$ *is an injective immersion.*

(c) The canonical map $\widehat{\alpha}_X$ *is an embedding, i.e.* $\mathcal{O}(X)$ *is a submanifold of* M *diffeomorphic to* $\mathbf{G}/\mathbf{H}_X$*, if and only if* $\widehat{\alpha}_X$ *is proper, cf. [139]. In this case, the tangent space of* $\mathcal{O}(X)$ *at* $Y = X\cdot G$ *is given by*

$$\begin{aligned} \mathrm{T}_Y\mathcal{O}(X) &= \mathrm{D}\,\alpha_X(G)\,\mathrm{T}_G\mathbf{G} \\ &= \mathrm{D}\,\alpha_Y(\mathbb{1})\,\mathfrak{g} \\ &= \mathrm{D}\,\alpha_Y(\mathbb{1})\,\mathrm{Ad}_{G^{-1}}\,\mathfrak{p}_X, \end{aligned} \tag{75}$$

where $\mathfrak{p}_X$ *is any complementary subspace of* $\mathfrak{h}_X$*, i.e.* $\mathfrak{g} = \mathfrak{h}_X \oplus \mathfrak{p}_X$*.*

Proof. (a) The continuity of α_X implies that $\mathbf{H}_X = \alpha_X^{-1}(X)$ is closed.

(b) In order to see that $\widehat{\alpha}_X$ is an injective immersion, consider the identity $\alpha_X \circ r_G = \alpha\big(\alpha_X(\cdot), G\big)$ and thus

$$\mathrm{D}\,\alpha_X(G)\cdot gG = \mathrm{D}_1\,\alpha(X,G) \circ \mathrm{D}\,\alpha_X(\mathbb{1})\,g.$$

Therefore, $\mathrm{D}\,\alpha_X(\mathbb{1})\,g = 0$ implies

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\alpha_X\big(\exp(tg)\big) = 0$$

for all $t \in \mathbb{R}$ and hence $\ker\mathrm{D}\,\alpha_X(\mathbb{1}) \subset \mathfrak{h}_X$. As the inclusion $\mathfrak{h}_X \subset \ker\mathrm{D}\,\alpha_X(\mathbb{1})$ is obvious, we obtain

$\ker \mathrm{D}\,\alpha_X(\mathbb{1}) = \mathfrak{h}_X$. Moreover, let $\mathfrak{p}_X$ be any complementary subspace of $\mathfrak{h}_X$. Then, identifying $\mathfrak{p}_X$ with $\mathrm{T}_{[\mathbb{1}]}\mathbf{G}/\mathbf{H}_X$ yields $D\widehat{\alpha}_X\big([\mathbb{1}]\big) = \mathrm{D}\,\alpha_X(\mathbb{1})\big|_{\mathfrak{p}}$, cf. Theorem III.7. Thus, $D\widehat{\alpha}_X\big([\mathbb{1}]\big)$ is injective and the same holds for any other $[G] \in \mathbf{G}/\mathbf{H}_X$ by right multiplication $\widehat{r}_G$.

(c) The first part follows from a standard embedding criterion on immersed manifolds, cf. [86]. The first equality of Eqn. (75) is a straightforward consequence of the identity $\alpha_X = \widehat{\alpha}_X \circ \Pi_X$, where $\Pi_X : \mathbf{G} \to \mathbf{G}/\mathbf{H}_X$ denotes the canonical projection. The second one is obtained by $\alpha_Y = \alpha_X \circ l_G = \alpha_X \circ r_G \circ \mathrm{Ad}_G$, while the third one follows from $\mathbf{H}_Y = \mathrm{Ad}_{G^{-1}} \mathbf{H}_x$. For further details see also [21]. □

Corollary III.3. *Let $\alpha : M \times \mathbf{G} \to M$ be as in Theorem III.9 and let $X \in M$ be any point.*

(a) If $\mathbf{G}$ is compact then $\mathbf{G}/\mathbf{H}_X$ is diffeomorphic to $\mathcal{O}(X)$.

(b) If α is transitive then $\mathbf{G}/\mathbf{H}_X$ is diffeomorphic to M.

Proof. (a) This follows immediately from Theorem III.9(c) and the compactness of $\mathbf{G}$.

(b) Observe that transitivity of α implies surjectivity of $\mathrm{D}\,\alpha_X(G)$ and $\mathrm{D}\,\widehat{\alpha}_X\big([G]\big)$. Thus, Theorem III.9(b) yields the desired result, cf. [102]. □

This gives rise to the following definition. A manifold M is called a *homogeneous $\mathbf{G}$-space* or for short a *homogeneous space*, if there exists a transitive smooth Lie group action of $\mathbf{G}$ on M. In particular, any coset space $\mathbf{G}/\mathbf{H}$ can be regarded as a homogeneous space via the canonical action $([G'], G) \mapsto [G'G]$ for $[G'] \in \mathbf{G}/\mathbf{H}$ and $G \in \mathbf{G}$. Further results on homogeneous spaces, orbit spaces and principal $\mathbf{G}$-bundles can be found in [90, 98, 102]

Remark III.4. Note that by Theorem III.9 the orbit $\mathcal{O}(X)$ carries always a manifold structure the topology of which is equal or finer than the topology induced by M.

Reductive Homogeneous Spaces

Let M be homogeneous space with transitive Lie group action $\alpha : M \times \mathbf{G} \to M$ and let $\mathbf{H} := \mathbf{H}_X$ be the stabiliser subgroup of a fixed element $X \in M$. Next, we are interested in carrying over the Riemannian structure of $\mathbf{G}$ to M or, equivalently, to $\mathbf{G}/\mathbf{H}$. First, we need some further terminology. As most of the following terms are conveniently expressed via algebraic properties of the pair $(\mathbf{G}, \mathbf{H})$, we focus on the case $M = \mathbf{G}/\mathbf{H}$. Nevertheless, one could restate all result in terms of an abstract group action α on M.

A homogeneous space $\mathbf{G}/\mathbf{H}$ is *reductive*, if the Lie algebra $\mathfrak{h}$ of $\mathbf{H}$ has a complementary subspace $\mathfrak{p}$ in $\mathfrak{g}$ such that $\mathfrak{p}$ is $\mathrm{Ad}_{\mathbf{H}}$-invariant, i.e. $H\mathfrak{p}H^{-1} \subset \mathfrak{p}$ for all $H \in \mathbf{H}$.

A Riemannian metric $\langle\cdot|\cdot\rangle$ on $\mathbf{G}/\mathbf{H}$ is called $\mathbf{G}$-*invariant* if the mappings $\widehat{r}_G$ are isometries, i.e. if the identity

$$\langle \xi | \eta \rangle_{[G']} = \big\langle \mathrm{D}\,\widehat{r}_G([G'])\xi \big| \mathrm{D}\,\widehat{r}_G([G'])\eta \big\rangle_{[G'G]} \tag{76}$$

is satisfied for all $\xi, \eta \in \mathrm{T}_{[G']}\mathbf{G}/\mathbf{H}$ and $G, G' \in \mathbf{G}$. Moreover, a bilinear form $(\cdot|\cdot)$ on $\mathfrak{p}$ is called

(a) $\mathrm{Ad}_{\mathbf{H}}$-*invariant* if the identity

$$(p|p') = (\mathrm{Ad}_H\, p|\, \mathrm{Ad}_H\, p') \tag{77}$$

is satisfied for all $p, p' \in \mathfrak{p}$ and $H \in \mathbf{H}$.

(b) $\mathrm{ad}_{\mathfrak{h}}$-*invariant* if the identity

$$(\mathrm{ad}_h\, p|p') = -(p|\, \mathrm{ad}_h\, p') \tag{78}$$

is satisfied for all $p, p' \in \mathfrak{p}$ and $h \in \mathfrak{h}$.

Note that $\mathbf{G}/\mathbf{H}$ is reductive, if $\mathbf{G}$ has a bi-invariant metric, as one can choose $\mathfrak{p} := \mathfrak{h}^{\perp}$. Next, we give a generalisation of Proposition III.2 and Theorem III.4 to homogeneous spaces.

Proposition III.4. *Let $\mathbf{G}/\mathbf{H}$ be a homogeneous space with reductive decomposition $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{p}$. The following statements are equivalent:*

(a) There exists a $\mathbf{G}$-invariant metric $\langle\cdot|\cdot\rangle$ on $\mathbf{G}/\mathbf{H}$.

(b) There exists an $\mathrm{Ad}_{\mathbf{H}}$-invariant scalar product $(\cdot|\cdot)$ on $\mathfrak{p}$.

In addition, if $\mathbf{H}$ is connected then (a) and (b) are equivalent to

(c) There exists a $\mathrm{ad}_{\mathfrak{h}}$-invariant scalar product $(\cdot|\cdot)$ on $\mathfrak{p}$.

Proof. Cf. [94] and Proposition III.2. □

Theorem III.10. *Let $\mathbf{G}/\mathbf{H}$ be a homogeneous space with reductive decomposition $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{p}$. Then $\mathbf{G}/\mathbf{H}$ admits a $\mathbf{G}$-invariant metric if and only if the closure of $\mathrm{Ad}_{\mathbf{H}}|_{\mathfrak{p}} := \{\mathrm{Ad}_H : \mathfrak{p} \to \mathfrak{p} \mid H \in \mathbf{H}\}$ is compact in $GL(\mathfrak{p})$.*

Proof. Cf. [84]. □

Remark III.5. (a) As a special case, Theorem III.10 implies the existence of bi-invariant metrics on compact Lie groups, cf. Theorem III.4 and [84].

(b) Replacing $\mathfrak{p}$ by the quotient space $\mathfrak{g}/\mathfrak{h}$, allows to state Theorem III.10 without referring to any reductive decomposition $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{p}$ of $\mathfrak{g}$, cf. [84]. Moreover, it can be shown that any homogeneous space $\mathbf{G}/\mathbf{H}$ which admits a $\mathbf{G}$-invariant metric is reductive, cf. [103].

Theorem III.10 can easily be rephrased for an arbitrary homogeneous $\mathbf{G}$-space M with transitive group action $\alpha : M \times \mathbf{G} \to M$, by choosing $\mathbf{H} := \mathbf{H}_X$ with $X \in M$. Note however, for orbits $M := \mathcal{O}(X)$ embedded in some larger Riemannian manifold N, the invariant metric given by Theorem III.10 does in general not coincide with the induced metric. This gives rise to the following definition.

A manifold M is called a *Riemannian homogeneous* $\mathbf{G}$-space or for short *Riemannian homogeneous space*, if M is a homogeneous $\mathbf{G}$-space with α-*invariant metric*, which is to say that the mappings $\alpha_G : M \to M$, $\alpha_G(X) := X \cdot G$ are isometries of M for all $G \in \mathbf{G}$, i.e.

$$\langle \xi | \eta \rangle_X = \big\langle \mathrm{D}\,\alpha_G(X)\xi \big| \mathrm{D}\,\alpha_G(X)\eta \big\rangle_{X \cdot G} \tag{79}$$

for all $\xi, \eta \in \mathrm{T}_X M$ and $G \in \mathbf{G}$.

Proposition III.5. *(a) Any homogeneous space of the form* $\mathbf{G}/\mathbf{H}$ *with a* $\mathbf{G}$*-invariant metric is a Riemannian homogeneous space.*

(b) Any Riemannian homogeneous space is isometric to a homogeneous space of the form $\mathbf{G}/\mathbf{H}$ *with a* $\mathbf{G}$*-invariant metric.*

Proof. This follows straightforwardly from the previous definitions and Corollary III.3(b). □

Naturally Reductive Homogeneous Spaces and Geodesics

Characterising the Riemannian connection of a homogeneous space and its geodesics are in general advanced issues which we do not want to address here, cf. [94] and the references therein e.g. [104]. However, there are two cases – see (a) and (b) below – which are easy to handle. A homogeneous space $\mathbf{G}/\mathbf{H}$ is called

(a) *naturally reductive* if it is reductive with complementary space $\mathfrak{p}$ and $\mathrm{Ad}_{\mathbf{H}}$-invariant scalar product $(\cdot|\cdot)$ on $\mathfrak{p}$ such that the identity

$$(P\,\mathrm{ad}_g\, h|k) = -(h|P\,\mathrm{ad}_g\, k) \tag{80}$$

is satisfied for all $g, h, k \in \mathfrak{p}$, where P denotes the projection onto $\mathfrak{p}$ along $\mathfrak{h}$.

(b) *Cartan-like* if it is reductive with complementary space $\mathfrak{p}$ and $\mathrm{Ad}_{\mathbf{H}}$-invariant scalar product $(\cdot|\cdot)$ on $\mathfrak{p}$ such that the commutator relations

$$[\mathfrak{h}, \mathfrak{h}] \subset \mathfrak{h}, \quad [\mathfrak{h}, \mathfrak{p}] \subset \mathfrak{p} \quad \text{and} \quad [\mathfrak{p}, \mathfrak{p}] \subset \mathfrak{h}. \tag{81}$$

are satisfied.

Remark III.6. If, in definition (a), the complementary space $\mathfrak{p}$ can be chosen as the orthogonal complement of $\mathfrak{h}$ with respect to some $\mathrm{Ad}_{\mathbf{G}}$-invariant scalar product $(\cdot|\cdot)$ on $\mathfrak{g}$, then condition (80) reduces to

$$(\mathrm{ad}_g\, h|k) = -(h|\,\mathrm{ad}_g\, k) \tag{82}$$

for all $g, h, k \in \mathfrak{p}$.

Lemma III.2. *Any Cartan-like Riemannian homogeneous space* $\mathbf{G}/\mathbf{H}$ *is naturally reductive.*

Proof. By the commutator relation $[\mathfrak{p}, \mathfrak{p}] \subset \mathfrak{h}$, we have $P\,\mathrm{ad}_g\, h = 0$ for all $g, h \in \mathfrak{p}$. Thus Eqn. (80) is obviously satisfied. □

Theorem III.11 (Coset Version)**.** *Let* $\mathbf{G}/\mathbf{H}$ *be naturally reductive. Then* $\mathbf{G}/\mathbf{H}$ *is Riemannian homogeneous space such that all geodesics through* $[G] \in \mathbf{G}/\mathbf{H}$ *are of the form*

$$t \mapsto \big[G \exp(t\,\mathrm{Ad}_{G^{-1}}\, p)\big] = \big[\exp(tp)G\big] \tag{83}$$

with $p \in \mathfrak{p}$.

Proof. Obviously, $\mathbf{G}/\mathbf{H}$ is Riemannian homogeneous space by Proposition III.4. For a proof for Eqn.(83) we refer to [94, 99]. □

The above result can be restated for an arbitrary naturally reductive Riemannian homogeneous $\mathbf{G}$-space.

Theorem III.12 (Orbit Version)**.** *Let* M *be a homogeneous* $\mathbf{G}$*-space with transitive group action* $\alpha : M \times \mathbf{G} \to M$*. Assume that* $\mathbf{G}/\mathbf{H}_X$ *is naturally reductive with decomposition* $\mathfrak{g} = \mathfrak{h}_X \oplus \mathfrak{p}_X$*. Then* M *is a Riemannian homogeneous* $\mathbf{G}$*-space such that all geodesics through* $Y = X \cdot G \in M$ *are of the form*

$$t \mapsto Y \cdot \exp(t\,\mathrm{Ad}_{G^{-1}}\, p) \tag{84}$$

with $p \in \mathfrak{p}_X$.

Proof. The result is a straightforward consequence of Theorem III.11. □

Thus, on *naturally reductive* spaces, the Riemannian exponential map is particularly simple to compute. By taking the basic framework of Ref. [101] further to discuss geodesics, Figure 4 illustrates that only in naturally reductive homogeneous spaces the geodesics project from the group $\mathbf{G}$ to give geodesics on $\mathbf{G}/\mathbf{H}$. Therefore, in this sense, in naturally reductive homogeneous spaces projection and exponentiation of tangent vectors commute. However, on reductive homogeneous spaces that are *not* naturally reductive, the problem is considerably more involved. A necessary and sufficient condition for $t \mapsto [G\exp(tg)]$ being a geodesic in $\mathbf{G}/\mathbf{H}$ can be found in [94, 105].

On the other hand, for numerical purposes it is often enough and even advisible to approximate the Riemannian exponential map by another computationally more efficient local parametrisation. Here, the map

$$\mathfrak{p} \ni p \mapsto \Pi \circ l_G \circ \exp(\mathrm{Ad}_{G^{-1}}\, p) \tag{85}$$

might be a natural candidate, even if it fails to give the exact Riemannian exponential map. These issues are subject to current research, and recent details can be found in [24, 106]. Figure 4 also shows how in reductive homogeneous spaces that are no longer naturally reductive, the projected geodesic still provides a first-order approximation to the geodesic generated by the projection of the tangent vector.

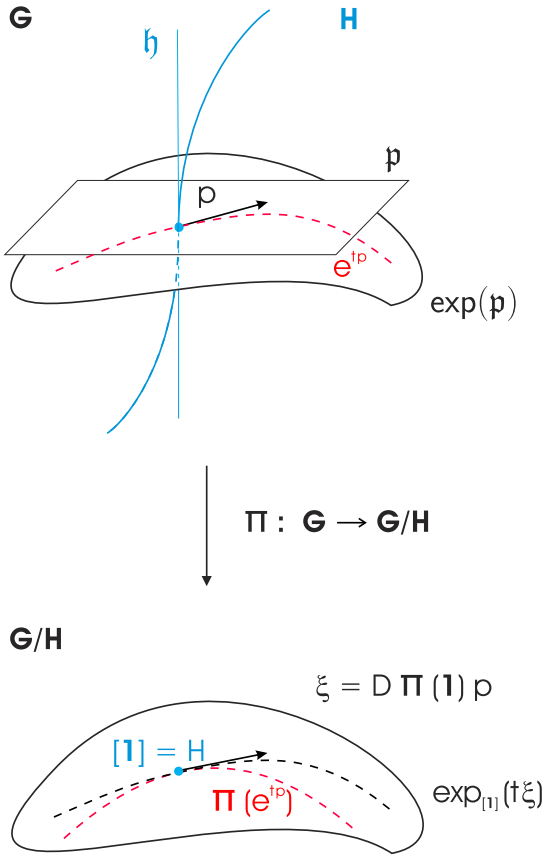

Figure 4: (Colour online) Geodesics in reductive homogeneous spaces $\mathbf{G}/\mathbf{H}$. The tangent vector $p \in \mathfrak{p}$ projects to the tangent vector ξ at the coset $[\mathbb{1}] = \mathbf{H}$. Note that only in *naturally* reductive homogeneous spaces the geodesic in $\mathbf{G}$ generated by p projects onto $\mathbf{G}/\mathbf{H}$ such that it coincides with the geodesic of the projected tangent vector in the sense $\Pi(e^{tp}) = \exp_{[\mathbb{1}]}(t\xi)$. In reductive homogeneous spaces that are *not* naturally reductive, the projection yields only a first-order approximation at $[\mathbb{1}] = \mathbf{H}$ as shown in the lower part, where $\Pi(e^{tp}) \neq \exp_{[\mathbb{1}]}(t\xi)$.

Adjoint Orbits

A prime example for naturally reductive homogeneous spaces is provided by the adjoint action of a compact Lie group—a scenario which is of major interest in the forthcoming applications. Therefore, we summarise the previous results for the particular case of adjoint orbits. Note that the adjoint action given by $(X, G) \mapsto \mathrm{Ad}_G X := GXG^{-1}$ is a *left* action. However, all previous statements and formulas remain valid *mutatis mutandis*, e.g., right cosets have to be replaced by left cosets, etc.

Corollary III.4. *Let $\mathbf{G}$ be a Lie group with Lie algebra $\mathfrak{g}$ and let $\mathbf{K} \subset \mathbf{G}$ be a compact subgroup with Lie algebra $\mathfrak{k}$ and bi-invariant metric $\langle\cdot|\cdot\rangle$. Moreover, let $\alpha : \mathfrak{g}\times\mathbf{K} \to \mathfrak{g}$, $(X, K) \mapsto \mathrm{Ad}_K X := KXK^{-1}$ be the adjoint action of $\mathbf{K}$ on $\mathfrak{g}$ and denote by $\alpha_X : \mathbf{K} \to \mathfrak{g}$ the map $K \mapsto \mathrm{Ad}_K X$. Then the following assertions hold*

(a) The stabiliser group $\mathbf{H} := \mathbf{H}_X$ of X is a closed subgroup of $\mathbf{K}$.

(b) The coset space $\mathbf{K}/\mathbf{H}$ is diffeomorphic to the adjoint orbit $\mathcal{O}(X) := \{\mathrm{Ad}_K X \mid K \in \mathbf{K}\}$ of X. In particular, the map $\widehat{\alpha}_X : \mathbf{K}/\mathbf{H} \to \mathcal{O}(X)$, $[K] \mapsto \mathrm{Ad}_K X$ is a well-defined diffeomorphism satisfying the commutative diagram

$$\begin{array}{ccc} \mathbf{K} & \xrightarrow{\alpha_X} & \mathcal{O}(X) \subset \mathfrak{g} \\ {\scriptstyle\Pi}\downarrow & \nearrow_{\widehat{\alpha}_X} & \\ \mathbf{K}/\mathbf{H} & & \end{array} \tag{86}$$

(c) Let $\mathfrak{h} := \mathfrak{h}_X$ denote the Lie algebra of $\mathbf{H}$ and $\mathfrak{p}$ be any complementary space to $\mathfrak{h}$ in $\mathfrak{k}$, then $\mathrm{D}\,\alpha(\mathbb{1}) = -\mathrm{ad}_X$ is a surjective homomorphism with $\ker \mathrm{ad}_X = \mathfrak{h}$ and commutative diagram

$$\begin{array}{ccc} \mathfrak{g} & \xrightarrow{\mathrm{D}\,\alpha_X(\mathbb{1})} & \mathrm{T}_X\mathcal{O}(X) \subset \mathfrak{g} \\ {\scriptstyle \mathrm{D}\,\Pi(\mathbb{1})}\downarrow & \nearrow_{\mathrm{D}\,\widehat{\alpha}_X([\mathbb{1}])} & \\ \mathfrak{p} \cong \mathfrak{k}/\mathfrak{h}\,; & & \end{array} \tag{87}$$

Moreover, the tangent space of $\mathcal{O}(X)$ at $Y = \mathrm{Ad}_K X$ is given by

$$\mathrm{T}_Y\mathcal{O}(X) = \mathrm{ad}_Y\,\mathfrak{k} = \mathrm{ad}_Y(\mathrm{Ad}_{K^{-1}}\,\mathfrak{p}).$$

(d) $\mathcal{O}(X) \cong \mathbf{K}/\mathbf{H}$ is naturally reductive. More precisely, $\mathfrak{p} := \mathfrak{h}^\perp$ yields a naturally reductive decomposition of $\mathfrak{k}$ with $\mathrm{Ad}_\mathbf{H}$-invariant scalar product on $\mathfrak{p}$ is given by the restriction of $\langle\cdot|\cdot\rangle$.

(e) There is a well-defined α-invariant metric on $\mathcal{O}(X)$ given by

$$\langle\xi|\eta\rangle_{\mathrm{Ad}_K X} := \langle p_\xi | p_\eta\rangle \tag{88}$$

with $\xi = \mathrm{ad}_Y\left(\mathrm{Ad}_K p_\xi\right)$, $\eta = \mathrm{ad}_Y\left(\mathrm{Ad}_K p_\eta\right)$ and $p_\xi, p_\eta \in \mathfrak{p}$.

(f) All geodesics through $Y = \mathrm{Ad}_K X \in \mathcal{O}(X)$ with respect to the metric given in part (e) are of the form

$$t \mapsto \mathrm{Ad}_{\exp(t\,\mathrm{Ad}_K\,p)}\,Y \tag{89}$$

with $p \in \mathfrak{p}$.

Proof. Part (a) and (b) follow immediately from Theorem III.9 and Corollary III.3.

(c) For $k \in \mathfrak{k}$ we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{Ad}_{\exp(tk)}\,X\Big|_{t=0} = -\,\mathrm{ad}_X\,k$$

and thus $\mathrm{D}\,\alpha(\mathbb{1}) = -\,\mathrm{ad}_X$. All other statemants are again consequences of Theorem III.9.

(d) First, observe that the bi-invariance of $\langle\cdot|\cdot\rangle$ implies that $\mathfrak{k} = \mathfrak{h} \oplus \mathfrak{p}$ with $\mathfrak{p} := \mathfrak{h}^\perp$ is reductive. Now, let P denote the orthogonal projection onto $\mathfrak{p}$. In turn, the bi-invariance of $\langle\cdot|\cdot\rangle$ yields

$$\langle P \operatorname{ad}_g h|k\rangle = \langle \operatorname{ad}_g h|k\rangle = -\langle h|\operatorname{ad}_g k\rangle = -\langle h|P\operatorname{ad}_g k\rangle$$

for all $g, h, k \in \mathfrak{p}$, cf. Proposition III.2. Therefore, $\mathcal{O}(X) \cong \mathbf{K}/\mathbf{H}$ is naturally reductive.

(e) Let $Y \in \mathcal{O}(X)$ and $\widetilde{K} \in \mathbf{K}$. A straightforward calculation using the identities $\mathrm{D}\,\alpha_{\widetilde{K}}(Y)\xi = \operatorname{Ad}_{\widetilde{K}}\xi$ for $\xi \in \mathrm{T}_Y\mathcal{O}(X)$ and $\operatorname{Ad}_{\widetilde{K}}(\operatorname{ad}_Y k) = \operatorname{ad}_{\operatorname{Ad}_{\widetilde{K}} Y}(\operatorname{Ad}_{\widetilde{K}} k)$ for all $k \in \mathfrak{k}$ yields the required invariance.

Part (f) follows immediately from Theorem III.12 and the identity $\mathfrak{h}_Y = \operatorname{Ad}_K \mathfrak{h}_X$ for $Y = \operatorname{Ad}_K X$ which implies $\mathfrak{h}_Y^\perp = \operatorname{Ad}_K \mathfrak{p}$. □

Gradient Flows on Riemannian Homogeneous Spaces

Applying the previous results on gradient flows to quality functions $\widehat{f}$ on Riemannian homogeneous spaces $\mathbf{G}/\mathbf{H}$, we obtain by the $\mathbf{G}$-invariance of the Riemannian metric—similar to (52)—the gradient equality

$$\operatorname{grad} \widehat{f}\,([G]) = \mathrm{D}\widehat{r}_G([\mathbb{1}])\operatorname{grad}(\widehat{f} \circ \widehat{r}_G)\,([\mathbb{1}]) \tag{90}$$

for all $G \in \mathbf{G}$, where $\widehat{r}_G$ denotes the mapping $[G'] \mapsto [G'G]$. Therefore, the gradient of $\widehat{f}$ is completely determined by

$$G \mapsto \operatorname{grad}(\widehat{f} \circ \widehat{r}_G)([\mathbb{1}]) \in \mathfrak{p}. \tag{91}$$

However, Eqn. (91) does not induce a mapping from $\mathbf{G}/\mathbf{H}$ to $\mathfrak{p}$, as in general

$$\operatorname{grad}(\widehat{f} \circ \widehat{r}_G)([\mathbb{1}]) \neq \operatorname{grad}(\widehat{f} \circ \widehat{r}_{HG})([\mathbb{1}])$$

for $H \in \mathbf{H} \setminus \{\mathbb{1}\}$. The corresponding gradient system reads

$$[\dot{G}] = \mathrm{D}\widehat{r}_G([\mathbb{1}])\operatorname{grad}(\widehat{f} \circ \widehat{r}_G)([\mathbb{1}]). \tag{92}$$

For analysing the asymptotic behaviour of Eqn. (92), Subsection III A again provides the appropriate tools. For instance, if $\mathbf{G}/\mathbf{H}$ is compact we have.

Corollary III.5. *Let $\mathbf{G}/\mathbf{H}$ be a compact Riemannian homogeneous space and let $\widehat{f} : \mathbf{G}/\mathbf{H} \to \mathbb{R}$ be real analytic. Then any solution of Eqn. (92) converges to a critical point of $\widehat{f}$ for $t \to +\infty$.*

Proof. This follows immediately from Proposition III.1 and Theorem III.1 as a Riemannian homogeneous space constitutes always a real analytic Riemannian manifold, cf. [93, 98]. □

Finally, we return to our starting point and ask for the relation between (53) and (92) in the case of an $\mathbf{H}$-equivariant quality function f. Here, f induces a quality function $\widehat{f}$ on $\mathbf{G}/\mathbf{H}$ via

$$\widehat{f}([G]) := f(G) \tag{93}$$

for all $G \in \mathbf{G}$. Moreover, assume $\mathbf{G}$ carries a bi-invariant metric $\langle\cdot|\cdot\rangle$ and $\mathbf{G}/\mathbf{H}$ is a homogeneous space with reductive decomposition $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{p}$ and $\mathfrak{p} := \mathfrak{h}^\perp$. This implies that the restriction of $\langle\cdot|\cdot\rangle$ to $\mathfrak{p}\times\mathfrak{p}$ is $\operatorname{Ad}_{\mathbf{H}}$-invariant. Now, the identity $\widehat{f} \circ \Pi = f$ yields

$$\mathrm{D}\,\widehat{f}([G]) \cdot \mathrm{D}\,\Pi(G) = \mathrm{D}\,f(G) \quad \text{for all } G \in \mathbf{G}. \tag{94}$$

and hence

$$\big(\mathrm{D}\,\Pi(G)\big)^* \operatorname{grad}\widehat{f}([G]) = \operatorname{grad} f(G) \tag{95}$$

for all $G \in \mathbf{G}$, where Π denotes the canonical projection and $(\cdot)^*$ the adjoint mapping. By identifying $\mathfrak{p}$ with the tangent space of $\mathbf{G}/\mathbf{H}$ at $[\mathbb{1}]$, the map $\mathrm{D}\,\Pi(\mathbb{1})$ represents the orthogonal projector $h + p \mapsto p$ for $h \in \mathfrak{h}$ and $p \in \mathfrak{p}$. Hence, we obtain

$$\mathrm{D}\,\Pi(\mathbb{1})\big(\mathrm{D}\,\Pi(\mathbb{1})\big)^* = \mathrm{id}_{\mathfrak{p}}. \tag{96}$$

In the same way, using the identity $\Pi \circ r_G = \widehat{r}_G \circ \Pi$, one shows

$$\mathrm{D}\,\Pi(G)\big(\mathrm{D}\,\Pi(G)\big)^* = \mathrm{id}_{\mathrm{T}_{[G]}\mathbf{G}/\mathbf{H}} \tag{97}$$

for all $G \in \mathbf{G}$. Thus (95) yields

$$\operatorname{grad}\widehat{f}([G]) = \mathrm{D}\,\Pi(G)\operatorname{grad} f(G) \tag{98}$$

for all $G \in \mathbf{G}$. Therefore, we have proven the following result:

Theorem III.13. *Suppose $\mathbf{G}/\mathbf{H}$ satisfies the above assumptions and $f : \mathbf{G} \to \mathbb{R}$ is a $\mathbf{H}$-equivariant quality function with induced quality function $\widehat{f} : \mathbf{G}/\mathbf{H} \to \mathbb{R}$. Then the canonical projection of the gradient flow of Eqn. (53) onto $\mathbf{G}/\mathbf{H}$ yields the gradient flow of Eqn. (92), i.e., if $G(t)$ is a solution of Eqn. (53) then $\Pi(G(t))$ is one of Eqn. (92) .*

D. Examples

Often practically relevant quality functions take the form of a linear functional restricted to an adjoint orbit $\mathcal{O}(X)$. For instance, in quantum dynamics the unitary orbit $\mathcal{O}(A) := \{UAU^\dagger \mid U \in SU(N)\}$ of an initial state A plays a central role, because it defines the largest reachability set under closed Hamiltonian dynamics. Then the set of feasible expectation values is such a linear map, since it is the projection onto an observable C in the sense of a Hilbert-Schmidt scalar product. These expectation values can be generalised to arbitrary complex

square matrices $A, C \in \mathbb{C}^{N\times N}$ such as to coincide with the *C-numerical range*

$$W(C,A) := \{\mathrm{tr}(C^\dagger U A U^\dagger) \mid U \in SU(N)\} \quad . \tag{99}$$

As C-numerical ranges are well established in the mathematical literature [107, 108], in the sequel we will adopt the notation.

Note that finding the maximum absolute value, i.e., the *C-numerical radius*

$$r(C,A) := \max_{U\in SU(N)} |\,\mathrm{tr}\{C^\dagger U A U^\dagger\}| \tag{100}$$

is straightforward for Hermitian A, C (it amounts to sorting the respective eigenvalues, cf. Corollary III.8), while for arbitrary complex A, C there is no general analytical solution. Moreover, when restricting to local unitary operations $K \in SU_{\mathrm{loc}}(2^n) := SU(2)^{\otimes n}$, the maximisation task becomes non-trivial even for Hermitian A, C [109, 110].

Therefore, we now illustrate the previous theory by gradient flows on the entire unitary group $SU(2^n)$, on the local unitary group $SU(2)^{\otimes n}$ as well as their adjoint orbits.

Worked Example: $SU(N)$

Recall that $SU(N)$ is a compact connected Lie group of real dimension N^2-1. Its Lie algebra, i.e. its tangent space at the identity is given by set $\mathfrak{su}(N)$ of all skew-Hermitian matrices Ω with $\mathrm{tr}\,\Omega = 0$, i.e.,

$$\mathfrak{su}(N) := \{\Omega \in \mathbb{C}^{N\times N} \mid \Omega^\dagger = -\Omega,\, \mathrm{tr}\,\Omega = 0\}. \tag{101}$$

So elements $\Omega \in \mathfrak{su}(N)$ relate to Hamiltonians H via $\Omega = iH$. The tangent space at an arbitrary element $U \in SU(N)$ is

$$\mathrm{T}_U SU(N) = \mathfrak{su}(N)U = \{\Omega U \,|\, \Omega \in \mathfrak{su}(N)\}, \tag{102}$$

cf. Eqn. (45). Moreover, let $SU(N)$ be endowed with the bi-invariant Riemannian metric

$$\langle \Omega U | \Xi U\rangle_U := \mathrm{tr}(\Omega^\dagger \Xi), \tag{103}$$

defined on the tangent spaces $\mathrm{T}_U SU(N)$, cf. Eqn (47).

Now set

$$\begin{aligned} g : SU(N) &\to \mathbb{C}^{N\times N}, \quad g(U) := C^\dagger U A U^\dagger \\ f : SU(N) &\to \mathbb{R}, \quad f(U) := \mathrm{Re}\,\mathrm{tr}\{C^\dagger U A U^\dagger\} \end{aligned}$$

For computing the tangent map of g, we exploit the fact that $SU(N)$ is an embedded submanifold of $\mathbb{C}^{N\times N}$. Therefore, the tangent map is obtained by restricting the ordinary Fréchet derivative $\mathrm{D}g(U)$ to the tangent space $\mathrm{T}_U SU(N)$, cf. Appendix A. Thus, by applying the product rule, one easily finds

$$\begin{aligned} \mathrm{D}g(U)(\Omega U) &= C^\dagger \Omega U A U^\dagger + C^\dagger U A (\Omega U)^\dagger \\ &= C^\dagger \Omega U A U^\dagger - C^\dagger U A U^\dagger \Omega \,. \end{aligned} \tag{104}$$

Now, the chain rule as well as the short-hand notations $\tilde{A} := UAU^\dagger$ and $[\cdot,\cdot]_S$ to denote the skew-Hermitian part of the commutator $[\cdot,\cdot]$ give

$$\begin{aligned} \mathrm{D}\, f(U)(\Omega U) &= \mathrm{D}(\mathrm{Re}\ \mathrm{tr})(g(U)) \circ \mathrm{D}(g(U))(\Omega U) \\ &= \mathrm{Re}\,\mathrm{tr}\{C^\dagger \Omega \tilde{A} - C^\dagger \tilde{A}\Omega\} \\ &= \mathrm{Re}\,\mathrm{tr}\{[\tilde{A}, C^\dagger]\Omega\} \\ &= \mathrm{tr}\{[\tilde{A}, C^\dagger]_S \Omega\} \\ &= \langle [\tilde{A}, C]_S^\dagger | \Omega\rangle \\ &= \langle [\tilde{A}, C^\dagger]_S^\dagger U | \Omega U\rangle \,, \end{aligned} \tag{105}$$

where the last identity explicitly invokes the right-invariance of the Riemannian metric on $SU(N)$, cf. Eqn. (103). Now, identifying the above expression with

$$Df(U)(\Omega U) = \langle \mathrm{grad}\, f(U) | \Omega U\rangle \tag{106}$$

one gets the gradient vector field

$$\mathrm{grad}\, f(U) = [\tilde{A}, C^\dagger]_S^\dagger\, U \tag{107}$$

and thus the gradient system

$$\dot{U} = \mathrm{grad}\, f(U) = -[\tilde{A}, C^\dagger]_S\, U \quad . \tag{108}$$

By making use of the Riemannian exponential, see Eqns. (9) and (11), we finally arrive at the discretisation

$$U_{k+1} = e^{-\alpha_k [U_k A U_k^\dagger, C]_S}\, U_k, \tag{109}$$

where $\alpha_k \geq 0$ denotes an appropriate step size.

Gradient Flows on $SU(N)$

Consider a fully controllable system on $SU(N)$ such as

$$(\Sigma) \qquad \dot{U}(t) = -\mathrm{i}\Big(H_d + \sum_{\substack{k=1 \\ \alpha\in\{x,y\}}}^{n} u_k(t) H_{k,\alpha}\Big) U(t) \tag{110}$$

with $H_d := \sum_{k<l} J_{kl}\sigma_{k,z}\sigma_{l,z}$, $H_{k,\alpha} := \sigma_{k,\alpha}$, $\alpha \in \{x,y\}$ and connected spin-spin coupling topology, cf. [27]. Here and in the sequel,

$$\sigma_x := \begin{bmatrix} 0 & \mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix}, \quad \sigma_y := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_z := \begin{bmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{bmatrix} \tag{111}$$

denote the Pauli matrices, which form an orthogonal basis of $\mathfrak{su}(2)$. Moreover, $\sigma_{k,\alpha}$, $\alpha \in \{x,y,z\}$ is defined by

$$\sigma_{k,\alpha} := \mathbb{1}_2 \otimes \cdots \otimes \mathbb{1}_2 \otimes \sigma_\alpha \otimes \mathbb{1}_2 \otimes \cdots \otimes \mathbb{1}_2, \tag{112}$$

where the term σ_α appears in the k-th position of the Kronecker product and $\mathbb{1}_2$ denotes the 2×2-identity matrix.

As (Σ) is fully controllable by assumption, the entire group $SU(N)$ can be generated by evolutions under the

Hamiltonian of the system plus the available controls. If A is an initial density operator or a matrix collecting its signal-relevant terms, the orbit of the canonical semigroup action of (Σ) on A yields in this case the entire unitary orbit $\mathcal{O}(A) = UAU^\dagger$. Recall its 'projection' on some observable C (or its signal-relevant terms) forms the *C-numerical range of A* (see Eqn.100).

In this setting, there are two geometric optimisation tasks of particular practical relevance as they determine maximal signal intensity in coherent spectroscopy [26].

(a) *Find all points on the unitary orbit of A that minimise the Euclidean distance to C.*

(b) *Find all points on the unitary orbit of A that minimise the angle to the 1-dimensional, complex subspace spanned by C.*

Clearly, the distance

$$\begin{aligned} \|UAU^\dagger - C\|_2^2 = \\ \|A\|_2^2 + \|C\|_2^2 - 2\mathrm{Re}\{\mathrm{tr}\left(C^\dagger UAU^\dagger\right)\} \end{aligned} \tag{113}$$

is minimal if the overlap $\mathrm{Re}\{\mathrm{tr}\left(C^\dagger UAU^\dagger\right)\}$ is maximal. Moreover, making use of the definition of the angle between 1-dimensional complex subspaces

$$\cos^2(\measuredangle\{UAU^\dagger, C\}) := \frac{|\,\mathrm{tr}\{C^\dagger UAU^\dagger\}|^2}{\|A\|_2^2 \cdot \|C\|_2^2}\,. \tag{114}$$

problem (b) is equivalent to maximising the function $|\mathrm{tr}(C^\dagger UAU^\dagger)|$. Its maximal value is the C-numerical radius of A (see Eqn. 99). Obviously, $r_C(A) \leq \|A\|_2 \cdot \|C\|_2$ with equality if and only if $UAU^\dagger$ and C are complex collinear for some $U \in SU(N)$. Note that the two tasks (a) and (b) are equivalent whenever the C-numerical range forms a circular disk in the complex plane (centred at the origin); conditions for circular symmetry have been characterised in [111].

Extending concepts of Brockett [16] from the orthogonal to the special unitary group [26, 27, 112], the above optimisation problems (a) and (b) can be treated by the previously presented gradient-flow methods, cf. also [21, 22].

For fixed matrices $A, C \in \mathbb{C}^{N\times N}$ define

$$f_1 \;:\; SU(N) \to \mathbb{R}, \quad f_1(U) := \mathrm{Re}\,\mathrm{tr}(C^\dagger UAU^\dagger) \tag{115}$$

and

$$f_2 \;:\; SU(N) \to \mathbb{R}, \quad f_2(U) := |\mathrm{tr}(C^\dagger UAU^\dagger)|^2. \tag{116}$$

Observe that the distance problem (a) is solved by maximising f_1, while the angle problem is solved for maximal f_2.

Now, the differential and the gradient of f_1 with respect to the bi-invariant Riemannian metric Eqn. (103) is precisely given by the previous example as

$$\begin{aligned} Df_1(U)(\Omega U) &= \mathrm{Re}\,\mathrm{tr}([UAU^\dagger, C^\dagger]\Omega), \\ \mathrm{grad}\, f_1(U) &= [UAU^\dagger, C^\dagger]_S^\dagger\, U. \end{aligned}$$

The differential and the gradient of f_2 can be obtained in the same manner as

$$\begin{aligned} Df_2(U)(\Omega U) &= \mathrm{tr}(C^\dagger UAU^\dagger)^* \cdot \mathrm{tr}([UAU^\dagger, C^\dagger]\Omega) \\ &\quad - \mathrm{tr}(C^\dagger UAU^\dagger) \cdot \mathrm{tr}([UAU^\dagger, C^\dagger]^\dagger\Omega), \\ \mathrm{grad}\, f_2(U) &= 2\big(f_2(U)^* \cdot [UAU^\dagger, C^\dagger]\big)_S^\dagger U. \end{aligned}$$

This yields the following result.

Theorem III.14. *The gradient systems of f_ν, $\nu = 1, 2$ with respect to the bi-invariant Riemannian metric (103) are given by*

$$\dot{U} \;=\; \Omega_\nu(U)U \tag{117}$$

with

$$\Omega_1(U) := [UAU^\dagger, C^\dagger]_S^\dagger \quad \textit{and} \quad \Omega_2(U) := 2\big(f_2(U)^* \cdot [UAU^\dagger, C^\dagger]\big)_S^\dagger. \tag{118}$$

respectively. Each solution of (117) converges to a respective critical point for $t \to +\infty$. Thereby, the critical points of f_ν are characterised by $\Omega_\nu(U) = 0$, $\nu = 1, 2$.

Proof: The above computations immediately yield Eqn. (117). As f_ν, $\nu = 1, 2$ are real analytic, the convergence of each solution to a critical point is guaranteed by Proposition III.1 and Theorem III.1, cf. [112]. □

An implementable numerical integration scheme for the above gradient systems making use of the Riemannian exponential, see Eqns. (9) and (11), is given by

$$U_{k+1}^{(\nu)} \;=\; \exp\left(\alpha_k^{(\nu)}\Omega_\nu(U_k^{(\nu)})\right) U_k^{(\nu)}, \quad U_0 = \mathbb{1}_N \tag{119}$$

A suitable choice of step sizes $\alpha_k^{(\nu)} > 0$ ensuring convergence can be found in [26, 27, 112]. Generically, it drives $U_k^{(\nu)}$ into final states attaining the maxima of the quality functions f_ν, $\nu = 1, 2$. However, there is no guarantee that the gradient flows always reach the *global* maxima. Standard numerical integration procedures such as *e.g.* the Euler method are not applicable here as they would not preserve unitarity.

Gradient Flows on the Local Subgroup $SU_{\mathrm{loc}}(2^n)$

The quality functions introduced in the previous subsection may be restricted to the subgroup of local action, i.e. to

$$SU_{\mathrm{loc}}(2^n) := \underbrace{SU(2) \otimes \cdots \otimes SU(2)}_{n\text{-times}} \subset SU(2^n). \tag{120}$$

The Lie subalgebra to $SU_{\mathrm{loc}}(2^n) \subset \mathfrak{su}(2^n)$ can be specified by

$$\mathfrak{su}_{\mathrm{loc}}(2^n) := \left\{ \sum_{j=1}^n \mathbb{1}_2 \otimes \cdots \otimes \mathbb{1}_2 \otimes \Omega_j \otimes \mathbb{1}_2 \otimes \cdots \otimes \mathbb{1}_2 \,\middle|\, \Omega_j \in \mathfrak{su}(2) \right\},$$

where the term $\Omega_j \in \mathfrak{su}(2)$ appears at the j-th position, cf. Eqn. (112). Therefore, the tangent space of $SU_{\rm loc}(2^n)$ at an arbitrary element U is given by

$$\mathrm{T}_U SU_{\rm loc}(2^n) = \{\Omega U \mid \Omega \in \mathfrak{su}_{\rm loc}(2^n)\}. \tag{121}$$

Finally, $SU_{\rm loc}(2^n)$ is endowed with the bi-invariant Riemannian metric induced by $SU(2^n)$, i.e.

$$\langle \Omega U, \Xi U\rangle_U := \mathrm{tr}(\Omega^\dagger \Xi) \tag{122}$$

for $\Omega U, \Xi U \in \mathrm{T}_U SU_{\rm loc}(2^n)$.

Lemma III.3. *Let $\mathbf{H} \subset GL(N,\mathbb{C})$ be any closed subgroup with Lie algebra $\mathfrak{h} \subset \mathfrak{gl}(N,\mathbb{C}) := \mathbb{C}^{N\times N}$. Moreover let $h_1,\dots,h_m$ be a real orthonormal basis of $\mathfrak{h}$ with respect to the real scalar product*

$$(g_1|g_2) := \mathrm{Re}\,\mathrm{tr}\,(g_1^\dagger g_2), \quad g_1, g_1 \in \mathbb{C}^{N\times N}, \tag{123}$$

i.e. $\mathrm{span}_{\mathbb{R}}\{h_1,\dots,h_m\} = \mathfrak{h}$ *and* $(h_i|h_j) = \delta_{ij}$.

(a) Then the orthogonal projection $P : C^{N\times N} \to C^{N\times N}$ onto $\mathfrak{h}$ is given by

$$g \mapsto Pg := \sum_{j=1}^m \mathrm{Re}\,\mathrm{tr}\{h_j^\dagger g\} h_j\ . \tag{124}$$

(b) The orthogonal projection $P^\perp : C^{N\times N} \to C^{N\times N}$ onto the orthogonal complement $\mathfrak{h}^\perp$ is given by

$$g \mapsto P^\perp g = g - Pg. \tag{125}$$

Proof: Both, (a) and (b) are basic and well-known facts from linear algebra. □

Remark III.7. For the unitary case, i.e. for $\mathfrak{h} \subset \mathfrak{su}(N)$, the real part in Eqn. (124) can be neglected and the projector P can be rewritten in the more convenient matrix form $\widehat{P}$ as

$$\widehat{P} := \sum_{j=1}^m \mathrm{vec}(h_j)\,\mathrm{vec}(h_j)^\dagger, \tag{126}$$

where the terms $\mathrm{vec}(h_j)\,\mathrm{vec}(h_j)^\dagger$ represent the rank-1 projectors $P_j = |h_j\rangle\langle h_j|$ in vec-notation.

Corollary III.6. *The orthogonal projection $P : C^{N\times N} \to C^{N\times N}$ onto $\mathfrak{su}_{\rm loc}(2^n)$ with respect to (123) is given by*

$$\begin{aligned} Pg := \frac{1}{2^n}\sum_{k=1}^n \Big(&\mathrm{Re}\,\big(\mathrm{tr}(g^\dagger X_k)\big) X_k + \mathrm{Re}\,\big(\mathrm{tr}(g^\dagger Y_k)\big) Y_k \\ &+ \mathrm{Re}\,\big(\mathrm{tr}(g^\dagger Z_k)\big) Z_k\Big), \end{aligned} \tag{127}$$

where X_k, Y_k and Z_k are defined by, cf. Eqn. (112)

$$X_k := \sigma_{k,x}, \quad Y := \sigma_{k,y}, \quad Z := \sigma_{k,z}. \tag{128}$$

Proof: This follows straightforwardly from the orthogonality of the set $\{X_k, Y_k, Z_k \mid k = 1,\dots,n\}$ and Lemma III.3. □

Theorem III.15. *Let $f_{\rm loc}$ be the restriction of (115) to $SU_{\rm loc}(2^n)$.*

(a) The gradient of $f_{\rm loc}$ with respect (122) and the corresponding gradient system are given by

$$\mathrm{grad}\, f_{\rm loc}(U) = P([C^\dagger, UAU^\dagger])U \tag{129}$$

and

$$\dot U = P([C^\dagger, UAU^\dagger])U, \tag{130}$$

respectively, where P denotes the orthogonal projection $P : \mathfrak{gl}(2^n,\mathbb{C}) \to \mathfrak{gl}(2^n,\mathbb{C})$ onto $\mathfrak{su}_{\rm loc}(2^n)$. More explicitly, (130) is equivalent to a system of n coupled equations

$$\dot U_k = \Omega_k U_k, \quad k = 1,\cdots,n \tag{131}$$

on $SU(2)$, where

$$\begin{aligned} \Omega_k = \frac{1}{2^n}\Big(&\mathrm{Re}(\mathrm{tr}([C^\dagger, UAU^\dagger]^\dagger X_k))X \\ &+ \mathrm{Re}(\mathrm{tr}([C^\dagger, UAU^\dagger]^\dagger Y_k))Y \\ &+ \mathrm{Re}(\mathrm{tr}([C^\dagger, UAU^\dagger]^\dagger Z_k))Z\Big). \end{aligned}$$

Each solution of (130) converges for $t \to \pm\infty$ to a critical point of $f_{\rm loc}$. Thereby, the critical points are characterised by

$$P([C^\dagger, UAU^\dagger]) = 0. \tag{132}$$

(b) The Hessian form $\mathrm{Hess} f_{\rm loc}(U)$ and the Hessian operator $\mathbf{Hess} f_{\rm loc}(U)$ of $f_{\rm loc}$ at U are given by

$$\begin{aligned} &\mathrm{Hess} f_{\rm loc}(U)(\Omega U, \Xi U) \ = \\ &= \ \frac{1}{2}\Big(\mathrm{Re}\,\big(\mathrm{tr}(\Omega^\dagger [C^\dagger, [\Xi, UAU^\dagger]])\big) \\ &\quad + \mathrm{Re}\,\big(\mathrm{tr}(\Omega^\dagger [UAU^\dagger, [\Xi, C^\dagger]])\big)\Big). \end{aligned} \tag{133}$$

and

$$\mathbf{Hess} f_{\rm loc}(U)\Omega U \ = \ \big(\mathbf{S}(U)\Omega\big)U, \tag{134}$$

respectively, with $\Omega \in \mathfrak{su}_{\rm loc}(2^n)$ and

$$\begin{aligned} &S(U)\Omega \ := \\ &\frac{1}{2}P\Big([C^\dagger, [\Omega, UAU^\dagger]] + [UAU^\dagger, [\Omega, C^\dagger]]\Big). \end{aligned} \tag{135}$$

(c) For all initial points $U_0 \in SU_{\rm loc}(2^n)$ the discretization scheme

$$U_{k+1} := \exp\Big(\alpha_k P\big([C^\dagger, U_k AU_k^\dagger]\big)\Big)\, U_k \tag{136}$$

with step size $\alpha_k :=$

$$\frac{\big\|P\big([C^\dagger, U_k AU_k^\dagger]\big)\big\|^2}{\big\|[C^\dagger, P\big([C^\dagger, U_k AU_k^\dagger]\big)]\big\| \cdot \big\|[P\big([C^\dagger, U_k AU_k^\dagger]\big), U_k AU_k^\dagger]\big\|} \tag{137}$$

converges to the set of critical points of $f_{\rm loc}$.

Proof: The subsequent arguments follow our conference report [113], which also contains a complete proof for the flow on the entire groups such as $SU(2^n)$.

(a) Since $SU_{\mathrm{loc}}(2^n)$ is a closed subgroup of $SU(2^n)$, it is also an embedded Lie subgroup and thus a submanifold of $SU(2^n)$, cf. Remark III.2. Therefore, the gradient of f_{loc} is well-defined by (36). Furthermore, by (55) and (118) we obtain

$$\begin{aligned}&\operatorname{grad} f_{\mathrm{loc}}(U) \;=\\ &=\; P(\operatorname{grad} f_1(U)) \;=\; P\big([UAU^\dagger, C^\dagger]^\dagger\big)U\\ &=\; P\big([C^\dagger, UAU^\dagger]\big)U,\end{aligned}$$

where the last equality follows from $P([UAU^\dagger, C^\dagger]^\dagger) = -P([UAU^\dagger, C^\dagger])$ and the skew-symmetry of the commutator. Moreover, Eqn. (131) is derived by Corollary III.6 and the identity

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\Big(U_1(t)\otimes\cdots\otimes U_n(t)\Big) \;=\\ &\left(\sum_{k=1}^{n} \mathbb{1}_2\otimes\cdots\otimes \dot{U}_k(t)U_k^{-1}(t)\otimes\cdots\otimes\mathbb{1}_2\right)\times\\ &\times\Big(U_1(t)\otimes\cdots\otimes U_n(t)\Big).\end{aligned}$$

Now, compactness of $SU_{\mathrm{loc}}(2^n)$ and real analyticity of f_{loc} imply that each solution converges to critical points for $t\to+\infty$, cf. Proposition III.1 and Theorem III.1.

(b) By (41), the Hessian of f_{loc} at U is determined by evaluating the second derivative of $\varphi := f\circ\gamma$ at $t=0$, where γ is any geodesics. This yields

$$\begin{aligned}\operatorname{Hess} f_{\mathrm{loc}}(U)(\Omega U,\Omega U) \;&:=\; \varphi''(0)\\ &=\; \operatorname{Re}\big(\operatorname{tr}(C^\dagger[\Omega,[\Omega,UAU^\dagger]])\big),\end{aligned}\tag{138}$$

for $\Omega\in\mathfrak{su}_{\mathrm{loc}}(2^n)$. The Hessian then is obtained from quadratic form (138) by a standard polarisation argument, i.e.

$$\begin{aligned}&\operatorname{Hess} f_{\mathrm{loc}}(U)(\Omega U,\Xi U) \;=\\ &=\; \frac{1}{2}\Big(\operatorname{Re}\big(\operatorname{tr}(C^\dagger[\Omega,[\Xi,UAU^\dagger]])\big)\\ &\quad+\operatorname{Re}\big(\operatorname{tr}(C^\dagger[\Xi,[\Omega,UAU^\dagger]])\big)\Big).\end{aligned}$$

Finally, by the identity $\operatorname{tr}[X,Y]Z = -\operatorname{tr} Y[X,Z]$ we conclude

$$\begin{aligned}&\operatorname{Hess} f_{\mathrm{loc}}(U)(\Omega U,\Xi U) \;=\\ &=\; \frac{1}{2}\Big(\operatorname{Re}\big(\operatorname{tr}(\Omega^\dagger[C^\dagger,[\Xi,UAU^\dagger]])\big)\\ &\quad+\operatorname{Re}\big(\operatorname{tr}(\Omega^\dagger[UAU^\dagger,[\Xi,C^\dagger]])\big)\Big).\end{aligned}$$

Therefore, the Hessian operator of f_{loc} at U is given by

$$\mathbf{Hess} f_{\mathrm{loc}}(U)\Omega U \;=\; \big(\mathbf{S}(U)\Omega\big)U\tag{139}$$

with $\Omega\in\mathfrak{su}_{\mathrm{loc}}(2^n)$ and

$$\begin{aligned}&\mathbf{S}(U)\Omega \;:=\\ &\frac{1}{2}P\Big([C^\dagger,[\Omega,UAU^\dagger]]+[UAU^\dagger,[\Omega,C^\dagger]]\Big).\end{aligned}\tag{140}$$

(c) Estimating the second derivative

$$\varphi''(t) = \operatorname{Re}\big(\operatorname{tr}([C^\dagger,\Omega][\Omega,\mathrm{e}^{t\Omega}UAU^\dagger\mathrm{e}^{-t\Omega}])\big)\tag{141}$$

for $\Omega := \operatorname{grad} f_{\mathrm{loc}}(U) = P([C^\dagger,UAU^\dagger])$ and $U\in SU_{\mathrm{loc}}(2^n)$ yields

$$\begin{aligned}|\varphi''(t)| \;&\le\; \Big\|[C^\dagger,\Omega]\Big\|\cdot\Big\|[\Omega,\mathrm{e}^{t\Omega}UAU^\dagger\mathrm{e}^{-t\Omega}]\Big\|\\ &=\; \Big\|[C^\dagger,\Omega]\Big\|\cdot\Big\|[\Omega,UAU^\dagger]\Big\|.\end{aligned}$$

Therefore, we get the estimate

$$\begin{aligned}&\max_{t\ge0}\Big|\frac{\mathrm{d}^2}{\mathrm{d}t^2}f_{\mathrm{loc}}(\exp_U(\Omega t))\Big|\\ &\le\; \Big\|[C^\dagger,\Omega]\Big\|\cdot\Big\|[\Omega,UAU^\dagger]\Big\|\end{aligned}$$

for $\Omega := \operatorname{grad} f_{\mathrm{loc}}(U)$. Now, a standard Lyapunov-type argument, similar to the proof of Theorem 3.3 in cf. [21], yields the desired result.

□

For similar discretisation schemes in different contexts or other intrinsic Riemannian methods see also [18, 21, 26, 114].

Double-Bracket Flows on Naturally Reductive Homogeneous Spaces

The well-known *double-bracket flows* have established themselves as useful tools for diagonalising matrices (usually real symmetric ones) as well as for sorting lists [16, 18, 21, 22]. Moreover, they relate to Hamiltonian integrateble systems [115, 116]. (Note again that in many-particle physics gradient flows were later introduced independently for diagonalising Hamiltonians [50, 51].) In summarising the most important results we show that double-bracket flows can be viewed as special cases of gradient flows on *naturally reductive* homogeneous spaces $\mathbf{G}/\mathbf{H}$ in terms of Sec. III C, where $\mathbf{H}$ is a stabiliser group, which is typically not normal. Then the homogeneous space $\mathbf{G}/\mathbf{H}$ does not constitute a group itself.

Let

$$\mathcal{O}(A) := \{UAU^\dagger | U\in SU(N)\}\tag{142}$$

denote the unitary orbit of some $A \in \mathbb{C}^{N\times N}$. Note that the *adjoint action* $(U, A) \mapsto \mathrm{Ad}_U A := UAU^\dagger$ of $SU(N)$ constitutes a *left* action on the Lie algebra $\mathfrak{g} := \mathbb{C}^{N\times N}$. However, this should not cause any confusion for the reader since the key result we refer to—Corollary III.4—was presented for left actions.

Let $C \in \mathbb{C}^{N\times N}$ be another complex matrix. For minimising the (squared) Euclidean distance $\|X - C\|_2^2$ between C and the unitary orbit of A we derive a gradient flow maximising the target function

$$\widehat{f}(X) := \mathrm{Re}\,\mathrm{tr}\{C^\dagger X\} \tag{143}$$

over $X \in \mathcal{O}(A)$. Clearly, this is but an alternative to tackling the problem by a gradient flow on the unitary group, since as in Sec. III C, we have the equivalence

$$\max_{X\in\mathcal{O}(A)} \widehat{f}(X) \;=\; \max_{U\in SU(N)} f(U) \tag{144}$$

for $f(U) := \mathrm{Re}\,\mathrm{tr}\{C^\dagger UAU^\dagger\}$.

Building upon Corollary III.4, we have the following facts: $\mathcal{O}(A)$ constitutes a compact and connected naturally reductive homogeneous space isomorphic to $SU(N)/\mathbf{H}$. Here,

$$\mathbf{H} := \{U \in SU(N) |\, \mathrm{Ad}_U A = A\} \tag{145}$$

denotes the stabiliser group of A. Recalling that the Lie algebra of $SU(N)$ is given by

$$\mathfrak{su}(N) := \{\Omega \in \mathbb{C}^{N\times N} \mid \Omega^\dagger = -\Omega\}, \tag{146}$$

we further obtain for the tangent space of $\mathcal{O}(A)$ at $X = \mathrm{Ad}_U A$ the from

$$T_X\mathcal{O}(A) = \big\{\mathrm{ad}_X \Omega \,\big|\, \Omega \in \mathfrak{su}(N)\big\} \tag{147}$$

with $\mathrm{ad}_X \Omega := [X, \Omega]$. Moreover, the kernel of $\mathrm{ad}_A : \mathfrak{su}(N) \to \mathfrak{g}$ reads

$$\mathfrak{h} = \{\Omega \in \mathfrak{su}(N) \mid [A, \Omega] = 0\}. \tag{148}$$

and forms the Lie subalgebra to $\mathbf{H}$. Now, by the standard Hilbert-Schmidt scalar product $(\Omega_1, \Omega_2) \mapsto \mathrm{tr}\{\Omega_1^\dagger\Omega_2\}$ on $\mathfrak{su}(N)$ one can define the ortho-complement to the above kernel as

$$\mathfrak{p} := \mathfrak{h}^\perp \quad . \tag{149}$$

This induces a unique decomposition of any skew-Hermitian matrix $\Omega = \Omega^\mathfrak{h} + \Omega^\mathfrak{p}$ with $\Omega^\mathfrak{h} \in \mathfrak{h}$ and $\Omega^\mathfrak{p} \in \mathfrak{p}$. Finally, we obtain an $\mathrm{Ad}_{SU(N)}$-invariant Riemannian metric on $\mathcal{O}(A)$ via

$$\langle \mathrm{ad}_X(\mathrm{Ad}_U \Omega_1) |\, \mathrm{ad}_X(\mathrm{Ad}_U \Omega_2)\rangle_X := \mathrm{tr}\{\Omega_1^{\mathfrak{p}\,\dagger}\Omega_2^\mathfrak{p}\} \tag{150}$$

for $X := \mathrm{Ad}_U A$, which is equivalent to saying

$$\langle \mathrm{ad}_X(\Omega_1) |\, \mathrm{ad}_X(\Omega_2)\rangle_X := \mathrm{tr}\{\Omega_1^{\mathfrak{p}_X\,\dagger}\Omega_2^{\mathfrak{p}_X}\} \tag{151}$$

with $\mathfrak{p}_X := \mathrm{Ad}_U \mathfrak{p}$. Now, the main results on double-bracket flows read as follows:

Theorem III.16. *Set* $\widehat{f} : \mathcal{O}(A) \to \mathbb{R}$, $\widehat{f}(X) := \mathrm{Re}\,\mathrm{tr}\{C^\dagger X\}$. *Then one finds*

(a) The gradient of $\widehat{f}$ *with respect to the Riemannian metric defined by Eqn.* (150) *is given by*

$$\mathrm{grad}\,\widehat{f}(X) = [X, [X, C^\dagger]_S], \tag{152}$$

where $[X, C^\dagger]_S$ *denotes the skew-Hermitian part of* $[X, C^\dagger]$.

(b) The gradient flow

$$\dot{X} = \mathrm{grad}\,\widehat{f}(X) = [X, [X, C^\dagger]_S] \tag{153}$$

defines an isospectral flow on $\mathcal{O}(A) \subset \mathfrak{g}$. *The solutions exist for all* $t \geq 0$ *and converge to a critical point* X_∞ *of* $\widehat{f}(X)$ *characterised by* $[X_\infty, C^\dagger]_S = 0$.

Proof. (A detailed proof for the real case can be found in [21]; for an abstract Lie algebraic version see also [18].)

(a) For $X = \mathrm{Ad}_U A$ and $\xi = \mathrm{ad}_X \Omega \in T_X\mathcal{O}(A)$ we obtain

$$\begin{aligned} \mathrm{D}\,\widehat{f}(X)\,\mathrm{ad}_X \Omega \;&=\; \frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{Re}\,\mathrm{tr}\{C^\dagger \mathrm{e}^{-t\Omega} X \mathrm{e}^{t\Omega}\}\Big|_{t=0} \\ &=\; \mathrm{Re}\,\mathrm{tr}\{C^\dagger\,\mathrm{ad}_X\,\Omega\}. \end{aligned}$$

Therefore, the gradient of $\widehat{f}$ has to satisfy

$$\mathrm{Re}\,\mathrm{tr}\{C^\dagger\,\mathrm{ad}_X\,\Omega^{\mathfrak{p}_X}\} = \langle \mathrm{grad}\,\widehat{f}(X) |\,\mathrm{ad}_X\,\Omega^{\mathfrak{p}_X}\rangle_X$$

for all $\Omega^{\mathfrak{p}_X} \in \mathfrak{p}_X$. Applying Eqn. (150) to $X = A$ gives

$$\mathrm{Re}\,\mathrm{tr}\{C^\dagger\,\mathrm{ad}_A\,\Omega^\mathfrak{p}\} = \mathrm{tr}\{\Gamma^{\mathfrak{p}\,\dagger}\Omega^\mathfrak{p}\}$$

for all $\Omega^\mathfrak{p} \in \mathfrak{p}$, where $\Gamma^\mathfrak{p}$ is defined by $\mathrm{grad}\,\widehat{f}(A) = \mathrm{ad}_A\,\Gamma^\mathfrak{p}$ with $\Gamma^\mathfrak{p} \in \mathfrak{p}$. Thus, we finally arrive at

$$\mathrm{tr}\{(\mathrm{ad}_{A^\dagger}\,C)_S^\dagger\Omega^\mathfrak{p}\} = \mathrm{tr}\{\Gamma^{\mathfrak{p}\,\dagger}\Omega^\mathfrak{p}\}$$

for all $\Omega^\mathfrak{p} \in \mathfrak{p}$, where $(\mathrm{ad}_{A^\dagger}\,C)_S$ denotes the skew-Hermitian part of $\mathrm{ad}_{A^\dagger}\,C$. Hence, $\Gamma^\mathfrak{p} = (\mathrm{ad}_{A^\dagger}\,C)_S^\mathfrak{p}$. Moreover, for $\Omega^\mathfrak{h} \in \mathfrak{h}$, we have

$$\begin{aligned} \mathrm{tr}\{(\mathrm{ad}_{A^\dagger}\,C)^\dagger\Omega^\mathfrak{h}\} \;&=\; -\,\mathrm{tr}\{\mathrm{ad}_A\,C^\dagger\Omega^\mathfrak{h}\} \\ &=\; \mathrm{tr}\{C^\dagger\,\mathrm{ad}_A\,\Omega^\mathfrak{h}\} \;=\; 0. \end{aligned}$$

Hence, $(\mathrm{ad}_{A^\dagger}\,C)_S \in \mathfrak{p}$ and therefore

$$\mathrm{grad}\,\widehat{f}(A) = \mathrm{ad}_A(\mathrm{ad}_{A^\dagger}\,C)_S = [A, [A, C^\dagger]_S].$$

The same arguments apply to $X = \mathrm{Ad}_U A$ and thus

$$\mathrm{grad}\,\widehat{f}(X) = [X, [X, C^\dagger]_S].$$

(b) Since Eqn. (152) evolves on the unitary orbit of A, the associated flow is isospectral by construction. The compactness of $\mathcal{O}(A)$ then implies that each solution

$X(t)$ of Eqn. (152) exists for all $t \geq 0$ and converges to the set of critical points cf. Proposition III.1. Moreover, from Theorem III.1 we derive that $X(t)$ converges actually to a single critical point X_∞ of $\widehat{f}$, i.e. to a point X_∞ which satisfies

$$[X_\infty, [X_\infty, C^\dagger]_S] = 0. \tag{154}$$

Since $[X_\infty, C^\dagger]_S \in \mathfrak{p}_{X_\infty}$, Eqn. (154) is equivalent to

$$[X_\infty, C^\dagger]_S = 0.$$

□

In order to obtain a numerical algorithm for maximising $\widehat{f}$ one can discretise the continuous-time gradient flow (152) as in the previous examples via

$$X_{k+1} = e^{-\alpha_k [X_k, C^\dagger]_S} \; X_k \; e^{\alpha_k [X_k, C^\dagger]_S} \tag{155}$$

with appropriate step sizes $\alpha_k > 0$. Note that Eqn. (155) heavily exploits the fact that the adjoint orbit $\mathcal{O}(A)$ constitutes a *naturally reductive homogeneous space* and thus the knowledge on its geodesics, cf. Corollary III.4.

Remark III.8. As an alternative to Eqn. 155, taking the standard Euler-type iteration

$$X_{k+1} = X_k + \alpha_k [X_k, [X_k, C^\dagger]_S] \tag{156}$$

does *not retain* the isospectral nature of the flow. Therefore, it should only be used as a computationally inexpensive, rough scheme in the neighbourhood of equilibrium points, if at all.

For A, C complex Hermitian (real symmetric) and the full unitary (or orthogonal) group or its respective orbit the gradient flow (152) is well understood, cf. Corollary III.8. However, for non-Hermitian A and C, the nature of the flow and in particular the critical points have not been analysed in depth, because the Hessian at critical points is difficult to come by. Even for A, C Hermitian, a full critical point analysis becomes non-trivial as soon as the flow is restricted to a closed and connected *subgroup* $\mathbf{K} \subset SU(N)$. Nevertheless, the techniques from Theorem III.16 can be taken over to establish a gradient flow and a respective gradient algorithm on the orbit $\mathcal{O}_\mathbf{K}$ in a straightforward manner.

Corollary III.7. *The gradient flow* (152) *restricts to the subgroup orbit* $\mathcal{O}_\mathbf{K}(A) := \{KAK^\dagger \mid K \in \mathbf{K} \subset SU(N)\}$ *by taking the respective orthogonal projection* $P_\mathfrak{k}$ *onto the subalgebra* $\mathfrak{k} \subset \mathfrak{su}(N)$ *of* $\mathbf{K}$ *instead of projecting onto the skew-Hermitian part, i.e.* $\dot{X} = [X, P_\mathfrak{k}[X, C^\dagger]]$.

The corresponding discrete integration scheme takes the form

$$X_{k+1} = e^{-\alpha_k P_\mathfrak{k}[X_k, C^\dagger]} \; X_k \; e^{\alpha_k P_\mathfrak{k}[X_k, C^\dagger]} \tag{157}$$

with appropriate step sizes $\alpha_k > 0$.

In view of unifying the interpretation of unitary networks, e.g., for the task of computing ground states of quantum mechanical Hamiltonians $H \equiv A$, the double-bracket flows for complex Hermitian A, C on the full unitary orbit $\mathcal{O}_u(A)$ as well as on the subgroup orbits $\mathcal{O}_\mathbf{K}(A)$ for different partitionings $\mathbf{K} := \{K \in SU(N_1) \otimes SU(N_2) \otimes \cdots \otimes SU(N_r) | \prod_{j=1}^r N_j = 2^n\}$ have shifted into focus [36]. Therefore, we have given the foundations for the the recursive schemes of Eqns. (155) and (157), which are listed in Table II as U1P and U1KP.

Finally, we summarise what is known about the nature of critical points for the real symmetric or complex Hermitian case. For a detailed discussion of the real symmetric case and the orthogonal group see e.g. [21].

Corollary III.8. *Let* C *and* A *be real symmetric or complex Hermitian and assume for simplicity that they show distinct eigenvalues in either case. Then one finds:*

(a) For A, C *real symmetric, define with respect to the special orthogonal group* $SO(N)$ *and* $Y \in \mathcal{O}_o(A) := \{OAO^\top | O \in SO(N)\}$ *a pair of target functions on the group and on the respective orbit by*

$$g(O) \;:=\; \mathrm{tr}\{C^\top OAO^\top\} \tag{158}$$

$$\widehat{g}(Y) \;:=\; \mathrm{tr}\{C^\top Y\} \quad . \tag{159}$$

Then the gradient flow

$$\dot{O} := \mathrm{grad}\; g(O) = [OAO^\top, C]\, O \tag{160}$$

shows $2^{(N-1)}N!$ *critical points, while the double-bracket flow*

$$\dot{Y} := \mathrm{grad}\; \widehat{g}(Y) = [Y, [Y, C]] \tag{161}$$

only shows $N!$ *equilibrium points.*

(b) For A, C *complex Hermitian, and* $X \in \mathcal{O}_u(A) := \{UAU^\dagger | U \in SU(N)\}$

$$f(U) \;:=\; \mathrm{tr}\{C^\dagger UAU^\dagger\} \tag{162}$$

$$\widehat{f}(X) \;:=\; \mathrm{tr}\{C^\dagger X\} \tag{163}$$

the gradient flow on the special unitary group $SU(N)$

$$\dot{U} := \mathrm{grad}\; f(U) = [UAU^\dagger, C]U \tag{164}$$

shows a continuum of critical points, while the double-bracket flow on the unitary orbit

$$\dot{X} := \mathrm{grad}\; \widehat{f}(X) = [X, [X, C]] \tag{165}$$

again shows only $N!$ *equilibrium points.*

(c) On the orbit, the respective target function has a unique global maximum which is given by the diagonalisation $\mathrm{diag}\,(\lambda_1, \ldots, \lambda_N)$, $\lambda_1 > \cdots > \lambda_N$

of A, if C is assumed to be diagonal of the form $C = \mathrm{diag}\,(\mu_1, \ldots, \mu_N)$, $\mu_1 > \cdots > \mu_N$. Moreover, the respective gradient flow converges to the unique global maximum for almost all initial values with an exponential bound on the rate.

Proof.
(a) and (b) The counting arguments follow immediately from the fact that in either case for C diagonal with distinct eigenvalues, the set of critical points $\mathcal{C}_\infty := \{X_\infty \in \mathcal{O}(A) | [X_\infty, C] = 0\}$ on the orthogonal or unitary *orbit* is given by $N!$ different diagonalisations of A and remains therefore invariant under conjugation by any permutation matrix.

Moreover, on the orthogonal group $O(N)$, the stabiliser group of A is given by

$$\{\mathrm{diag}\,(\pm 1, \pm 1, \ldots, \pm 1)\} \quad ,$$

which adds 2^N independent further degrees of freedom. Finally, restricting to $SO(N)$ we obtain $2^{N-1}N!$ critical points on the group level.

In contrast, for the unitary case $SU(N)$, the stabiliser group of A reads

$$\Big\{\mathrm{diag}\,(e^{\mathrm{i}\phi_1}, \ldots, e^{\mathrm{i}\phi_\nu}, \ldots, e^{\mathrm{i}\phi_N}) \;\Big|\; \sum_{\nu=1}^{N} \phi_\nu \in 2\pi\mathbb{Z},\ \phi_\nu \in \mathbb{R}\Big\},$$

which is always continuous.

(c) Since C is symmetric or Hermitian, we can assume without loss of generality that C is diagonal. Then, the critical point condition $[X_\infty, C]$ yields that the critical points of $\widehat{g}$ and, respectively, $\widehat{f}$ are given by the diagonalisations of A. Moreover, analysing the Hessian at critical points shows that *there is only one global maximum* in both cases and no local ones [21].

The exponential convergence of the gradient flows Eqns. (161) and (165) to the respective unique global maximum for almost all initial values is also established via the Hessian, i.e. by linearising the respective gradient flows at critical points [21]. □

Gradient Flows on Naturally Reductive Coset Spaces

More generally, let $\mathbf{G}$ be a Lie group with bi-invariant metric and let f be an *equivariant* quality function with respect to the closed subgroup $\mathbf{H}$, i.e. for all $H \in \mathbf{H}$ one has $f(G) = f(HG)$, so

$$f|_{\mathbf{H}G} = \text{constant}$$

for every $G \in \mathbf{G}$. Moreover, assume that $\mathbf{G}/\mathbf{H}$ is a naturally reductive coset space. Establishing a gradient method for the induced quality function $\widehat{f}$ on $\mathbf{G}/\mathbf{H}$ finally yields a recursion scheme which looks like the corresponding one on the group level, cf. Theorem III.11 and III.13. This, however, is not surprising, as the equivariance of f guarantees that the gradient (taken on the group level) at $G \in \mathbf{G}$ is orthogonal to the coset $\mathbf{H}G$. Thus, its 'pullback' to the Lie algebra automatically belongs to $\mathfrak{p}$ inducing a gradient flow on $\mathbf{G}/\mathbf{H}$.—This can be illustrated as follows.

With $\mathbf{G}/\mathbf{H}$ being naturally reductive, there is the reductive decomposition $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{p}$ with $\mathfrak{p} := \mathfrak{h}^\perp$, so any $\Omega \in \mathfrak{g}$ decomposes uniquely into $\Omega = \Omega^{\mathfrak{h}} + \Omega^{\mathfrak{p}}$. Then, by the equivariance of f one finds

$$\langle \mathrm{grad}\, f(G) | \Omega^{\mathfrak{h}} G \rangle = \mathrm{D}f(G)\, \Omega^{\mathfrak{h}} G = 0$$

for all $\Omega^{\mathfrak{h}} \in \mathfrak{h}$. Therefore, the 'pullback' of the gradient of f to $\mathfrak{g}$ satisfies $\mathrm{grad}\, f(G) G^{-1} \in \mathfrak{p}$. Furthermore, combining Eqns. (71), (98) and the identity

$$\mathrm{D}(\Pi \circ l_G)(\mathbb{1})\, \Omega = \mathrm{D}\Pi(G)\, G\Omega$$

for all $\Omega \in \mathfrak{g}$ (cf. Remark III.3) yields

$$\mathrm{grad}\, \widehat{f}([G]) = \mathrm{D}(\Pi \circ l_G)(\mathbb{1})\big(G^{-1}\, \mathrm{grad}\, f(G)\big). \qquad (166)$$

Thus, from Eqn. (83) we finally obtain

$$\exp_{[G]}\big(t\, \mathrm{grad}\, \widehat{f}([G])\big) = \big[\exp\big(t\, \mathrm{grad}\, f(G)\, G^{-1}\big)\, G\big]$$

for all $t \in \mathbb{R}$, where $\exp_{[G]}$ denotes the Riemannian exponential map at $[G]$, cf. Eqns. (10) and (11).—This precisely explains what we meant by the above statement 'looking like the one on the group level'.

Example Let $f : SU(2^n) \to \mathbb{R}$ be an arbitrary smooth function that is equivariant under local unitary operations of the n-fold tensor product $SU_{\mathrm{loc}}(2^n) := SU(2) \otimes \cdots \otimes SU(2)$. This includes, e.g., any measure of entanglement $\mu_E(U)$ that varies smoothly with U. Since by equivariant construction $\mathrm{grad}\, f|_{SU_{\mathrm{loc}}(2^n)} = 0$, we may consider the flow to $[\dot{U}] = \mathrm{grad}\, \widehat{f}([U])$ on the homogeneous space

$$\mathbf{G}/\mathbf{K} = SU(2^n)/SU_{\mathrm{loc}}(2^n) \quad ,$$

which is naturally reductive for all n and Cartan-like only for $n = 2$. This can be seen, because (i) $SU(2^n)$ carries a bi-invariant metric induced by the Killing form allowing to define $\mathfrak{p} := \mathfrak{k}^\perp$, which gives the reductive decomposition $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$, yet only for $n = 2$ one recovers the commutator inclusions $[\mathfrak{k}, \mathfrak{k}] \subseteq \mathfrak{k}$, $[\mathfrak{p}, \mathfrak{p}] \subseteq \mathfrak{k}$, and $[\mathfrak{k}, \mathfrak{p}] \subseteq \mathfrak{p}$; (ii) in any case, by Prop. III.4 there is an $\mathrm{Ad}_{\mathbf{K}}$-invariant scalar product on $\mathfrak{p}$; and (iii) Eqn. (80) is fulfilled for all $\{a, b, c\} \subseteq \mathfrak{p}$, as $\mathrm{tr}\{[a, b]^\dagger c\} = -\,\mathrm{tr}\{b^\dagger [a, c]\}$, cf. Remark III.6. — Clearly, this example generalises analogously to functions that are equivariant under actions of other partitionings of the full unitary group giving flows on the corresponding reductive homogeneous spaces

$$\mathbf{G}/\mathbf{K} = SU(N)/\big(SU(N_1) \otimes SU(N_2) \otimes \cdots \otimes SU(N_r)\big)$$

with $\prod_{j=1}^{r} N_j = N$.

Moreover, it is important to note that quality functions $\widehat{f}$ directly defined on $\mathbf{G}/\mathbf{H}$ (without resorting to equivariance) can be handled with the very same techniques as above, once the relation

$$\mathrm{grad}\, \widehat{f}([G]) = \mathrm{D}(\Pi \circ l_G)(\mathbb{1})\, \mathrm{Ad}_{G^{-1}}\, \Omega^{\mathfrak{p}}$$

for some $\Omega^{\mathfrak{p}} \in \mathfrak{p}$ is established, cf. Eqn. (166).

IV. APPLICATIONS TO QUANTUM CONTROL AND QUANTUM INFORMATION

A. A Geometric Measure of Pure-State Entanglement

The Euclidean distance of a pure state to the set $\mathcal{S}_{pp}$ of all pure product states may be seen as a geometric measure of *entanglement* [54, 117, 118]. Since $\mathcal{S}_{pp}$ coincides with the local unitary orbit

$$\mathcal{O}_{\mathrm{loc}}(yy^\dagger) := \{Uyy^\dagger U^\dagger \mid U \in SU_{\mathrm{loc}}(2^n)\} \tag{167}$$

of any pure product state $y \in \mathcal{S}_{pp}$, it relates to the following optimisation task

$$\Delta(x) \;:=\; \min_{U \in SU_{\mathrm{loc}}(2^n)} \|xx^\dagger - Uyy^\dagger U^\dagger\|^2, \tag{168}$$

where $x \in \mathbb{C}^{2^n}$ denotes a normalised pure state and $y \in \mathbb{C}^{2^n}$ a pure product state, e.g. $y = (1, 0, \ldots, 0)^\top = (e_1 \otimes \cdots \otimes e_1)$. This notation replaces $|x\rangle$ by x and $|x\rangle\langle x|$ by $xx^\dagger$ for the sake of convenient generalisation to higher order tensor products. Obviously, minimising (168) is equivalent to maximising the so-called *local transfer*

$$\max_{U \in SU_{\mathrm{loc}}(2^n)} \mathrm{Re}\left(\mathrm{tr}(xx^\dagger Uyy^\dagger U^\dagger)\right), \tag{169}$$

between $xx^\dagger$ and $yy^\dagger$. Further, since

$$\mathrm{tr}(xx^\dagger Uyy^\dagger U^\dagger) = \left|\mathrm{tr}(x^\dagger Uy)\right|^2$$

taking the real part in (169) is redundant.

Now, the techniques developed in Section III D match perfectly to tackle problem (169). Let $C := xx^\dagger$, $A := \mathrm{diag}\,(1, 0, \ldots, 0)$ and define the so-called local unitary transfer between C and A by the real-valued function

$$f_{\mathrm{loc}}(U) := \mathrm{tr}\,(CUAU^\dagger) \quad . \tag{170}$$

Then the gradient flow (130) or more precisely its discretisation (136) will generically solve (169). For explicit numerical results see Subsection IV B and [113, 119].

In general, neither an algebraic characterisation of the maximal value of f_{loc} nor the structure of its critical points is known, the major difficulty arising from the fact that U is restricted to $SU_{\mathrm{loc}}(2^n)$. As soon as U may be taken from the entire special unitary group, the solution is well-known: it is simply obtained by arranging the (real) eigenvalues of both A and C magnitude-wise in the same order [16, 21, 120, 121, 122].

B. Generalised Local Subgroup Optimisation

Bipartite Systems and Relations to Singular-Value Decompositions

An exceptional case, where the restricted problem (169) can be solved are bipartite pure systems. These systems are particularly simple in as much as the maxima of f_{loc} can be linked to the singular-value decomposition (SVD) of the matrices X and Y associated to x and y by $x := \mathrm{vec}\,X$ and $y := \mathrm{vec}\,Y$. Since these ideas readily extend to arbitrary finite dimensional bipartite systems, we generalise the formulation of Problem (169) thus leading to Eqn. (171), before going into multi-partite systems.

Proposition IV.1. *Let $X = V_X \Sigma_X W_X^\dagger$, $Y = V_Y \Sigma_Y W_Y^\dagger$ be singular value decompositions with $V_X, V_Y \in \mathrm{U}(N_1)$, $W_X, W_Y \in \mathrm{U}(N_2)$ and Σ_X, Σ_Y sorted by magnitude. Moreover, let $x := \mathrm{vec}\,X$ and $y := \mathrm{vec}\,Y$. Then the maximum value of the local transfer between $xx^\dagger$ and $yy^\dagger$ is bounded by*

$$\begin{aligned}\max_{U \in \mathrm{SU}(N_2) \otimes \mathrm{SU}(N_1)} &\mathrm{Re}\left(\mathrm{tr}(xx^\dagger Uyy^\dagger U^\dagger)\right) \\ &\leq (\mathrm{tr}\,\Sigma_X^\dagger \Sigma_Y)^2.\end{aligned} \tag{171}$$

Equality is actually achieved for $V_X, V_Y \in \mathrm{SU}(N_1)$, $W_X, W_Y \in \mathrm{SU}(N_2)$ and $U_* := (W_X^* \otimes V_X) \cdot (W_Y^\top \otimes V_Y^\dagger)$.

Proof: For $U := W \otimes V \in SU(N_2) \otimes SU(N_1)$ we obtain

$$\begin{aligned}&\mathrm{tr}(xx^\dagger Uyy^\dagger U^\dagger) \;= \\ &= \;\mathrm{tr}\left(xx^\dagger (W \otimes V) yy^\dagger (W^\dagger \otimes V^\dagger)\right) \\ &= \;\mathrm{tr}\left(xx^\dagger \,\mathrm{vec}(VYW^\top)\,\mathrm{vec}(VYW^\top)^\dagger\right) \\ &= \;\left|x^\dagger \,\mathrm{vec}(VYW^\top)\right|^2 \\ &= \;\left|\mathrm{tr}(X^\dagger VYW^\top)\right|^2.\end{aligned} \tag{172}$$

Here, we have used the identities

$$\begin{gathered}\mathrm{vec}(VYW) = (W^\top \otimes V)\,\mathrm{vec}\,Y, \\ (\mathrm{vec}\,X)^\dagger \,\mathrm{vec}\,Y = \mathrm{tr}\,X^\dagger Y\end{gathered}$$

for all $X, Y \in \mathbb{C}^{N_1 \times N_2}$. Now, (172) implies

$$\begin{aligned}&\max_{U \in \mathrm{SU}(N_2) \otimes \mathrm{SU}(N_1)} \mathrm{Re}\,\mathrm{tr}(xx^\dagger Uyy^\dagger U^\dagger) = \\ &= \max_{\substack{V \in \mathrm{SU}(N_1) \\ W \in \mathrm{SU}(N_2)}} \left|\mathrm{tr}(X^\dagger VYW^\top)\right|^2 \leq (\mathrm{tr}\,\Sigma_X^\dagger \Sigma_Y)^2,\end{aligned} \tag{173}$$

where the last inequality is due to von Neumann, cf. [107, 120]. If $V_X, V_Y \in \mathrm{SU}(N_1)$ and $W_X, W_Y \in \mathrm{SU}(N_2)$, equality is assumed in Eqn. (173) for

$$U_* := (W_Y W_X^\dagger)^\top \otimes V_X V_Y^\dagger = (W_X^* \otimes V_X) \cdot (W_Y^\top \otimes V_Y^\dagger).$$

□

Corollary IV.1. *Set $x := \mathrm{vec}\,A$ and $y := \mathrm{vec}\,C$. Then the maximum local transfer between $xx^\dagger$ and $yy^\dagger$ in the sense of Proposition IV.1 is bounded by*

$$||A||_C^2 := \max_{\substack{V \in \mathrm{U}(N_1) \\ W \in \mathrm{U}(N_2)}} \left|\mathrm{tr}(C^\dagger VAW^\dagger)\right|^2,$$

which is known as the C-spectral norm of A, cf. [108].

Note that in the context of finding maximal distances between *global unitary* orbits for the purpose of geometric discrimination of generic non-pure quantum states [123], results similar to [121, 122] show up, while here we treat *local unitary* orbits of pure bipartite states as explicit in Eqn. (171).

Multipartite Systems and Relations to Best Rank-1 Approximations of Higher Order Tensors

Proposition IV.1 has a straightforward generalisation to multipartite systems, which relates to best rank-1 approximations of higher order tensors. To outline this relation, we define the concept of a *generalised local subgroup*

$$SU_{\mathrm{loc}}(N_1,\ldots,N_r) := \mathrm{SU}(N_1)\otimes\cdots\otimes\mathrm{SU}(N_r). \quad (174)$$

of type $(N_1,\ldots,N_r)$ with $N_k\in\mathbb{N}$, $k=1,\ldots,r$. Thus, the associated general local subgroup optimisation Problem can be stated as follows.

Generalised Local Subgroup Optimisation Problem (GLSOP)

For $C, A\in\mathbb{C}^{N\times N}$ with $N := N_1\cdot N_2\cdots N_r$ find

$$\max_{U\in SU_{\mathrm{loc}}(N_1,\ldots,N_r)} \mathrm{Re}\left(\mathrm{tr}(CUAU^\dagger)\right). \quad (175)$$

To our knowledge, the GLSOP seems to be unsolved so far. To introduce higher order tensors, we have to fix some further notation. For simplicity, we regard a *tensor* of *order* $r\in\mathbb{N}$ as an array

$$X = (X_{i_1\ldots i_r})_{1\le i_1\le N_1,\ldots,1\le i_r\le N_r}$$

of size $N_1\times\cdots\times N_r$. The space of all $N_1\times\cdots\times N_r$-tensors is denoted by $\mathbb{C}^{N_1\times\cdots\times N_r}$. A natural scalar product for tensors of the same size is given by

$$\langle Y|X\rangle := \sum_{i_1\ldots i_r} Y^*_{i_1\ldots i_r}X_{i_1\ldots i_r} \quad . \quad (176)$$

Moreover, a tensor X is called a *rank-1 tensor* if there exist $x^k\in\mathbb{C}^{N_k}$, $k=1,\ldots,r$ such that

$$X = x^1\otimes_a x^2\otimes_a\cdots\otimes_a x^r, \quad (177)$$

where the $(i_1\ldots i_r)$-entry of the outer product $\otimes_a$ is defined by

$$(x^1\otimes_a x^2\otimes_a\cdots\otimes_a x^r)_{i_1\ldots i_r} := x^1_{i_1}\cdot x^2_{i_2}\cdots x^r_{i_r}.$$

Thus, the question of decomposing a given tensor by tensors of lower rank leads to the following fundamental approximation problem:

Best Rank-1 Approximation Problem (BRAP)

Let $\|\cdot\|$ denote the norm induced by scalar product (176). For $X\in\mathbb{C}^{N_1\times\cdots\times N_r}$ solve

$$\min_{\substack{C\in\mathbb{C},\|x^k\|=1\\ k=1,\ldots,r}} \|X - C\cdot x^1\otimes_a\cdots\otimes_a x^r\|^2. \quad (178)$$

Note that the above notation $\otimes_a$ is necessary to distinguish between two different types of outer products: the Kronecker product $\otimes$ (of column-vectors), which maps r-tuples of column-vectors to a column-vector of larger size, and the 'abstract' outer product $\otimes_a$, which maps r-tuples of column-vectors to arrays (= tensors) of order r. The relation between both is given by the canonical isomorphism $\mathfrak{vec}:\mathbb{C}^{N_1\times\cdots\times N_r}\to\mathbb{C}^N$ with $N := N_1\cdot N_2\ldots N_r$, which is uniquely determined by

$$x^1\otimes_a x^2\otimes_a\cdots\otimes_a x^r \mapsto x^1\otimes x^2\otimes\cdots\otimes x^r, \quad (179)$$

i.e. $\mathfrak{vec}$ assigns to each array $X\in\mathbb{C}^{N_1\times\cdots\times N_r}$ a column-vector in $\mathbb{C}^N$ by arranging the entries of X in a lexicographical order. With these notations at hand, the relation between GLSOP and BRAP can be stated as follows.

Theorem IV.1. *Let $X\in\mathbb{C}^{N_1\times\cdots\times N_r}$ be a tensor of order r and let $x := \mathfrak{vec}(X)\in\mathbb{C}^N$ with $N := N_1\cdot N_2\cdots N_r$. Then the* BRAP *is equivalent to the* GLSOP

$$\max_{U\in SU_{\mathrm{loc}}(N_1,\ldots,N_r)} \mathrm{Re}\left(\mathrm{tr}(xx^\dagger Uyy^\dagger U^\dagger)\right), \quad (180)$$

where $y\in\mathbb{C}^N$ can be any pure product state, e.g. $y = (1,0,\ldots,0)^\top = e_1\otimes\cdots\otimes e_1$. More precisely,

(a) If $U_1\otimes\cdots\otimes U_r$ is a solution of (180) then $x^k := U_ke_1$, $k=1,\ldots,r$ and $C := \langle X|x^1\otimes_a\cdots\otimes_a x^r\rangle$ solve (178).

(b) If $C\in\mathbb{C}$ and x^k, $k=1,\ldots,r$ solve (178) then any $U_1\otimes\cdots\otimes U_r$ with $x^k = U_ke_1$, $k=1,\ldots,r$ yields a solution of (180).

For proving Theorem IV.1 we need the following technical lemma.

Lemma IV.1. *The pair $(x^1\otimes_a\cdots\otimes_a x^r, C)$ solves (178) if and only if $x^1\otimes_a\cdots\otimes_a x^r$ is a maximum of*

$$\max_{\|z^k\|=1,k=1,\ldots,r} \left|\langle X|z^1\otimes_a\cdots\otimes_a z^r\rangle\right| \quad (181)$$

and $C = \langle X|x^1\otimes_a\cdots\otimes_a x^r\rangle$.

Proof: Consider the following identity

$$\begin{aligned}
&\|X - C\cdot z^1\otimes_a\cdots\otimes_a z^r\|^2\\
&= \|X\|^2 + |C|^2 - 2\,\mathrm{Re}\left(C^*\langle X|z^1\otimes_a\cdots\otimes_a z^r\rangle\right)\\
&= \|X\|^2 + \left|C - \langle X|z^1\otimes_a\cdots\otimes_a z^r\rangle\right|^2\\
&\quad -\left|\langle X|z^1\otimes_a\cdots\otimes_a z^r\rangle\right|^2.
\end{aligned}$$

Thus, we obtain

$$\begin{aligned}
&\min_{\substack{C\in\mathbb{C},\|z^k\|=1\\ k=1,\ldots,r}} \|X - C\cdot z^1\otimes_a\cdots\otimes_a z^r\|^2 =\\
&= \|X\|^2 - \max_{\substack{\|z^k\|=1\\ k=1,\ldots,r}} \left|\langle X|z^1\otimes_a\cdots\otimes_a z^r\rangle\right|^2.
\end{aligned}$$

This yields the desired result. □

Proof of Theorem IV.1: Let $y = e_1 \otimes \cdots \otimes e_1$. Then

$$\big(U_1 \otimes \cdots \otimes U_r\big)y = (U_1 e_1) \otimes \cdots \otimes (U_r e_1)$$

and thus

$$\begin{aligned}\mathrm{tr}(xx^\dagger U y y^\dagger U^\dagger) &= \mathrm{tr}(x^\dagger U y y^\dagger U^\dagger x)\\ &= \big|x^\dagger U y\big|^2\\ &= \big|\langle X|(U_1e_1)\otimes_a\cdots\otimes_a(U_re_1)\rangle\big|^2.\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}&\max_{U\in SU_{\mathrm{loc}}(N_1,\ldots,N_r)} \mathrm{Re}\,\big(\mathrm{tr}(xx^\dagger U y y^\dagger U^\dagger)\big) =\\ &= \max_{U\in SU_{\mathrm{loc}}(N_1,\ldots,N_r)} \big|\langle X|(U_1e_1)\otimes_a\cdots\otimes_a(U_re_1)\rangle\big|^2\\ &= \max_{\substack{\|z^k\|=1\\ k=1,\ldots,r}} \big|\langle X|z^1\otimes_a\cdots\otimes_a z^r\rangle\big|^2.\end{aligned}$$

and hence Lemma IV.1 implies (a) and (b). □

Remark IV.1.

1. The isomorphism $\mathfrak{vec}$ coincides 'almost' with the standard vec-operation on matrices for $r = 2$, more precisely $\mathfrak{vec}(X) = \mathrm{vec}(X^\top)$.

2. Since any phase factor can readily be absorbed into $x^1 \otimes_a \cdots \otimes_a x^r$, it is easy to show that

$$\begin{aligned}&\max_{\|x^k\|=1,k=1,\ldots,r} \big|\langle X|x^1\otimes_a\cdots\otimes_a x^r\rangle\big|\\ &= \max_{\|x^k\|=1,k=1,\ldots,r} \mathrm{Re}\,\big(\langle X|x^1\otimes_a\cdots\otimes_a x^r\rangle\big).\end{aligned}$$

   Therefore, maxima of the 'real-part-expression' on the right-hand side are always maxima of the 'absolute-value-term' on the left.

3. By replacing $yy^\dagger$ in (180) with an appropriate sum $\sum_{i=1}^l y_i y_i^\dagger$, the above ideas can be extended to best approximations of higher rank, i.e. to best approximations of the form

$$\min_{C_i\in\mathbb{C},\|x^{i,k}\|=1} \|X - \sum_{i=1}^{l} C_i\cdot x^{i,1}\otimes_a\cdots\otimes_a x^{i,r}\|^2,$$

   with $l \le \min\{N_1,\ldots,N_r\}$ and all $x^{i,1}\otimes_a\cdots\otimes_a x^{i,r}$ mutually orthogonal, cf. [124].

4. Unfortunately, an analogue of Proposition IV.1 involving the tensor SVD as defined in [125] does not hold for higher order tensors. Even the classical *Eckart-Young* Theorem, which asserts that the best rank-k approximation of a matrix is given by its truncated SVD, is false for higher order tensors, cf. [126].

Numerical Results

For comparing our gradient-flow approach to tensor-SVD techniques, here we focus on two examples that are well-established in the literature, since analytical solutions [127] as well as numerical results from semidefinite programming are known [54]. First, consider a pure 3-qubit state depending on a real parameter $s \in [0,1]$

$$|X(s)\rangle := \sqrt{s}|W\rangle + \sqrt{1-s}|V\rangle \quad , \tag{182}$$

where one defines

$$\begin{aligned}|W\rangle &:= \tfrac{1}{\sqrt{3}}\Big(|001\rangle + |010\rangle + |100\rangle\Big) \quad \text{and}\\ |V\rangle &:= \tfrac{1}{\sqrt{3}}\Big(|110\rangle + |101\rangle + |011\rangle\Big)\end{aligned}$$

with the usual short-hand notation of quantum information $|0\rangle := \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$, $|1\rangle := \left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$ and $|001\rangle := \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)\otimes\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)\otimes\left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$ etc. With these stipulations one finds the corresponding $2\times2\times2$ tensor representations for $|W\rangle$ and $|V\rangle$ to take the form

$$W_{(1,:,:)} = \tfrac{1}{\sqrt{3}}\begin{bmatrix}0&1\\1&0\end{bmatrix} \quad W_{(2,:,:)} = \tfrac{1}{\sqrt{3}}\begin{bmatrix}1&0\\0&0\end{bmatrix} \tag{183}$$

and

$$V_{(1,:,:)} = \tfrac{1}{\sqrt{3}}\begin{bmatrix}0&0\\0&1\end{bmatrix} \quad V_{(2,:,:)} = \tfrac{1}{\sqrt{3}}\begin{bmatrix}0&1\\1&0\end{bmatrix} . \tag{184}$$

Likewise, observe the pure 4-qubit-state

$$|\widehat{X}(s)\rangle := \sqrt{s}\,|GHZ'\rangle - \sqrt{1-s}\,|X^+\rangle\otimes|X^+\rangle \quad , \tag{185}$$

with the definitions

$$\begin{aligned}|\mathrm{GHZ}'\rangle &:= \tfrac{1}{\sqrt{2}}\Big(|0011\rangle + |1100\rangle\Big)\\ |X^+\rangle &:= \tfrac{1}{\sqrt{2}}\Big(|10\rangle + |01\rangle\Big) \quad .\end{aligned}$$

Consider the target function $f(K) = \mathrm{tr}\{C^\dagger K A K^\dagger\}$ with $C = \mathrm{diag}\,(1,0,0,\ldots,0)$ and $A := |\widehat{X}(s)\rangle\langle\widehat{X}(s)|$. As shown in Fig. 5 with the gradient flow restricted to the local unitaries $K \in SU_{\mathrm{loc}}(2^n)$ one obtains results perfectly matching the analytical solutions of [127] as well as the numerical ones from semidefinite programming ensuring global optimality—yet in drastically less CPU time as compared to [54], see Table I. Gradient flows are some 30 to 150 times faster in CPU time than semidefinite programming methods for the 3-qubit and 4-qubit example, respectively.

In the tensor-SVD algorithms [126] such as the higher-order power method (HOPM) or the higher-order orthogonal iteration (HOOI) as implemented in the MATLAB package [128], $N = 50$ to $N = 60$ iterations are required for quantitative agreement with the algebraically established results. In the 3-qubit example, all minimal distances are also reproduced correctly with $N = 5$ iterations–except for the limiting values s near 0 and near 1, for which the

Table I: CPU Times for Determining the Euclidean Distance to the Orbit of Separable Pure States as in Fig. 5

no. of qubits	semidefinite programming	*gradient flow on local unitaries*		tensor-SVD I (HOPM) [126, 128] higher-order power method		tensor-SVD II (HOOI) [126, 128] higher-order orthogonal iteration	
	CPU time [sec][a]	CPU time [sec][b]	speed-up	CPU time [sec][b]	speed-up	CPU time [sec][b]	speed-up
3	10.92	*0.30*	*36.4*	2.39	4.6	5.37	2.0
4	103.97	*0.71*	*147.0*	3.93	26.5	7.03	14.8

[a] Eisert *et al.* (processor with 2.2 GHz, 1 GB RAM) [54]
[b] average of 50 runs, Athlon XP1800+ (1.1 GHz, 512 MB RAM)

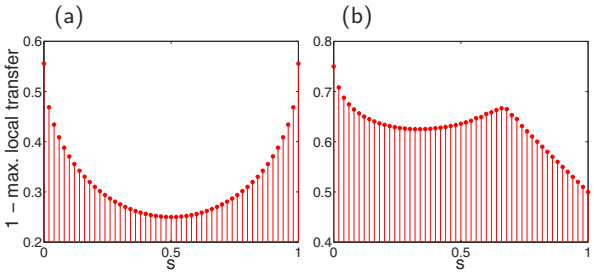

Figure 5: (Colour online) Numerical results by gradient flow on the local unitary group $\mathbf{K} = SU_{\rm loc}(2^n)$ determining (a) the Euclidean distance of the 3-qubit state $|X(s)\rangle = \sqrt{s}|W\rangle + \sqrt{1-s}|V\rangle$ (see Eqn. 182) to the nearest product state as a function of s; (b) the distance of 4-qubit state $|\widehat{X}(s)\rangle = \sqrt{s}|\mathrm{GHZ}'\rangle - \sqrt{1-\mathrm{s}}|\mathrm{X}^+\rangle \otimes |\mathrm{X}^+\rangle$(see Eqn. 185) to the nearest product state.

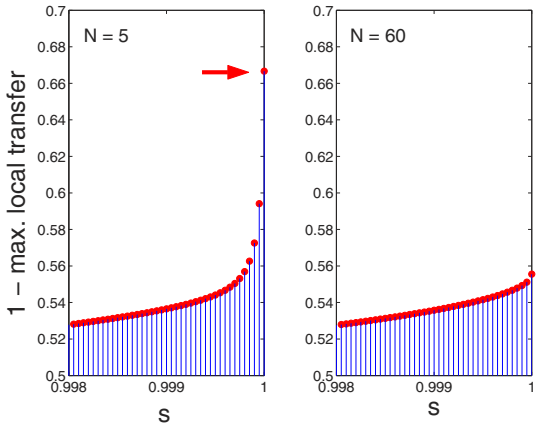

Figure 6: (Colour online) Tensor-SVD results for Euclidean distance of the 3-qubit state $|X(s)\rangle$ to the nearest product state as in Fig. 5(a). With the standard of $N = 5$ iterations, both methods (here shown for HOPM) give systematic errors as indicated by the arrow. $N = 60$ iterations are needed for quantitatively matching the well-established distance values. The high number of iterations required slows down the method as indicated in Tab. I.

minimal distances of $\Delta(|X(0)\rangle) = \Delta(|X(1)\rangle) = 2/3$ are obtained by either tensor method instead of the correct analytical value of 5/9, which requires $N = 60$ iterations as shown in Fig. 6. In the 4-qubit example, however, for $N = 5$ iterations, both tensor methods suffer from apparently random numerical instabilities, which only vanish when allowing for $N = 50$ iterations in either method. It is the considerably high number of iterations that makes the tensor methods substantially slower than our gradient-flow algorithm as shown in Tab. I.

Therefore, at least for lower order tensors, gradient flows provide an appealing alternative to tensor-SVD methods for best rank-1 approximations of higher order tensors. Moreover, one should take into account that the above gradient methods are developed to solve the GLSOP and thus a considerable speed-up can be expected by adjusting them to the local orbit $\mathcal{O}_{\rm loc}(yy^\dagger)$ of a pure product state. For a similar result obtained by an intrinsic Newton and conjugated gradient method see also [119]. — We anticipate that these numerical approaches will prove useful tools in tensor and rank aspects of entanglement and kinematics of qubit pairs as addressed, e.g., in Refs.[129, 130].

C. Locally Reversible Interaction Hamiltonians

Joint Local Reversibility

In a recent study [28], we have addressed the decision problem whether a time-independent (self-adjoint) Hamiltonian H normalised to $||H||_2 = 1$ generates a one-parameter unitary group $U(t) = \{e^{-itH}|t \in \mathbb{R}\}$ that is jointly invertible for all t by *local* unitary operations $K \in SU_{\rm loc}(2^n) = SU(2)^{\otimes n}$ in the sense

$$KHK^\dagger = -H \quad . \tag{186}$$

Apart from complete algebraic classification, in [28] we used that the question obviously finds an affirmative answer, if there is an element $K \in SU_{\rm loc}(2^n)$ such that

$$||KHK^\dagger + H||_2 = 0 \quad , \tag{187}$$

which amounts to *minimising* the transfer function

$$f(K) \;=\; \mathrm{Re}\ \mathrm{tr}\{HKHK^\dagger\} \quad . \tag{188}$$

With P denoting the projector onto $\mathfrak{k}$, i.e. the Lie algebra of $\mathbf{K} = SU_{\rm loc}(2^n)$, we therefore used the gradient flow

$$\dot{K} \;=\; -\,\mathrm{grad}\, f(K) = -P\big([KHK^\dagger, H]\big)\; K \quad (189)$$

as an other application of Theorem III.15. If (due to normalisation) Re $\mathrm{tr}\{HKHK^\dagger\} = -1$ can be reached, the interaction Hamiltonian is locally reversible.

Remark IV.2. There is an interesting relation to *local* C-numerical ranges as described in detail in Refs. [109, 110]: if the local C-numerical range

$$W_{\rm loc}(H,H) := \{\mathrm{tr}(HKHK^{-1}) | K \in \mathbf{K}\} = [-1;+1]$$

then the interaction Hamiltonian H is locally reversible. The references also establish the interconnection to local C-numerical ranges of circular symmetry and multi-quantum interaction components transforming like irreducible spherical spin tensors.

In Fig. 7, we give some examples: e.g., the Heisenberg ZZ interaction in a cyclic four-qubit coupling topology is locally reversible, while in the cyclic three-qubit topology or for the isotropic XXX interaction it is not. Thus numerical tests provide convenient answers in problems where an algebraic assessment becomes more tedious than in these examples, which are fully understood on algebraic grounds [28].

Pointwise Local Reversibility

In [28] we also generalised the above problem to the question, whether for a fixed $\tau \in \mathbb{R}$ there is a pair $K_1, K_2 \in \mathbf{K} = SU_{\rm loc}(2^n)$ so that

$$K_1 e^{-i\tau H} K_2 \;=\; e^{+i\tau H} \quad (190)$$

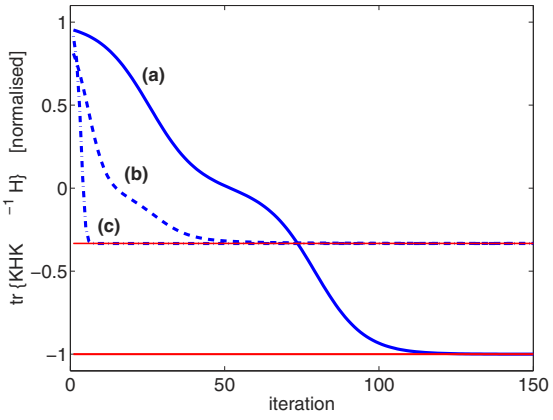

Figure 7: (Colour online) Gradient-flow driven local reversion of different Heisenberg interaction Hamiltonians: (a) the ZZ interaction on a cyclic four-qubit topology C_4 can be locally reversed, (c) the ZZ interaction on a cyclic three-qubit topology C_3 cannot be reversed locally, (c) nor the XXX isotropic interaction between two qubits.

which upon setting $A := e^{-i\tau H}$ and $C := e^{+i\tau H}$ is equivalent to

$$||K_1 A K_2 - C||_2 \;=\; 0 \quad . \quad (191)$$

Thus one may choose a gradient flow to *minimise*

$$f(K_1, K_2) := -\tfrac{1}{2^n}\,\mathrm{Re}\;\mathrm{tr}\{C^\dagger K_1 A K_2\} \quad (192)$$

by the coupled system

$$\dot{K}_1 \;=\; \mathrm{grad}\, f(K_1) = P(K_1 A K_2 C^\dagger)\; K_1 \quad (193)$$
$$\dot{K}_2 \;=\; \mathrm{grad}\, f(K_2) = P(K_2 C^\dagger K_1 A)\; K_2 \quad . \quad (194)$$

So if $f(K_1, K_2) = -1$ can be reached, then $U(t) = e^{-i\tau H}$ is locally reversible at time $t = \tau$. See Fig. 8 for examples comparing pointwise and universal local reversibility.

D. Constrained Optimisation in Quantum Control: Intrinsic vs Penalty Approach

In practical quantum control, one may face the problem to maximise a quality function f on the reachable set of a quantum system under *additional* state space contraints. For instance, find the maximal unitary transfer from matrix (state) A to C subject to leaving another state E invariant (provided A and E do not share the same stabiliser group). Another variant amounts to optimising the contrast between the transfer from A to C and the transfer from A to D; so the task is to maximise the transfer from A to C subject to suppressing the transfer from A to D.

For tackling those types of problems, we address two basically different approaches—a purely intrinsic one and a combined method joining intrinsic and penalty-type techniques. Both methods will be briefly illustrated for

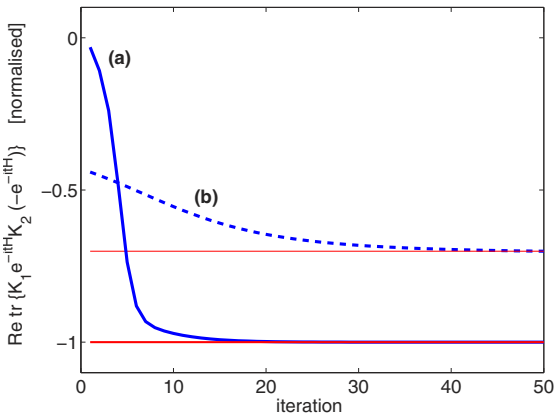

Figure 8: (Colour online) Gradient-flow driven local inversion of exponential of the Hamiltonian $H = \frac{1}{2}\,(\sigma_z \otimes \mathbb{1} + \mathbb{1} \otimes \sigma_z + \sigma_z \otimes \sigma_z)$ and $U(\tau) := \mathrm{e}^{-i\frac{\pi}{4}\,H}$ (a) by a gradient flow with K_1 and K_2 (b) by a gradient flow with $K_1 = K_2^\dagger =: K$.

Summary: General Gradient Algorithm for Steepest Ascent on Riemannian Manifolds

Requirements: Riemannian manifold M, e.g. Lie group $\mathbf{G}$ with (bi-invariant) metric $\langle\cdot|\cdot\rangle$ or its group orbits; smooth target function $f : M \to \mathbb{R}$; associated gradient system $\dot{X} = \operatorname{grad} f(X)$.

Input: initial state $X(0) \in M$, parameters for target function.

Output: sequence of iterative pairs $\{\big(X_k, f(X_k)\big)\}$ approximating critical points X_* and their critical values $f(X_*)$.

Initialisation: If possible, generate generic initial state X_0, e.g. for compact Lie groups pick random $G_0 \in \mathbf{G}$ according to Haar measure (for $SU(N)$ see [131]) and set $X_0 := G_0 \cdot X(0)$, otherwise identify $X_0 := X(0)$; calculate $f(X_0)$, $\operatorname{grad} f(X_0)$, and step size α_0 according to Section III.

Recursion:
while $k = 0, 1, 2, \ldots, k_{\text{limit}}$ **and** $\alpha_k > \alpha_{\text{threshold}} > 0$ **do**

1: iterate $X_{k+1} = \exp_{X_k}\big(\alpha_k \operatorname{grad} f(X_k)\big)$ (see collection of examples in Tab. II).

2: calculate $f(X_{k+1})$.

3: update step size α_{k+1} according to Section III.

4: go to step 1.

end

Figure 9: Summerising scheme for steepest-ascent gradient flows on Riemannian manifolds. For related methods, like conjugate gradients, Jacobi- or Newton-type schemes, step (1) has to be modified in a straight-forward way according to Sec. II, for details see Refs. [19, 61, 62]. If the dynamic stepsize selection of Sec. III is too costly CPU-timewise, one may start out with constant stepsizes, and halve them whenever $\big(f(X_{k+1}) - f(X_k)\big) \leq 0$, cf. Armijo's rule. In cases, where local extrema exist (see Sec. III), make sure to run with a sufficient number of generic initial conditions.

the problem of maximising the transfer from A to C while leaving E invariant, i.e.

$$\max_{U \in U(N)} |\operatorname{tr}\{UAU^\dagger C^\dagger\}| \quad \text{subject to } UEU^\dagger = E. \tag{195}$$

It is straightforward to see that the stabiliser group

$$\mathbf{K}_E := \{K \in U(N) \,|\, KEK^\dagger = E\} \tag{196}$$

of E forms a compact connected Lie subgroup of $U(N)$. Differentiating the identity $\mathrm{e}^{tk} E \mathrm{e}^{-tk} = E$ for $t = 0$ yields its Lie algebra

$$\mathfrak{k}_E := \{k \in \mathfrak{u}(N) \,|\, \operatorname{ad}_k(E) \equiv [k, E] = 0\}. \tag{197}$$

By the Jacobi identity

$$[[A, B], E] + [[B, E], A] + [[E, A], B] = 0 \tag{198}$$

for all $A, B, E \in \mathbb{C}^{N\times N}$, one can easiliy verify that $\mathfrak{k}_E$ is a Lie subalgebra of $\mathfrak{u}(N)$. Moreover, from the compactness of $\mathbf{K}_E$ we conclude that not only does $\mathfrak{k}_E$ generate the stabiliser group $\mathbf{K}_E$ in the sense $\langle \exp(\mathfrak{k}_E)\rangle = \mathbf{K}_E$, but also does so in the stronger sense $\exp(\mathfrak{k}_E) = \mathbf{K}_E$.

A set of generators of $\mathfrak{k}_E$ may constructively be found by solving a system of homogeneous linear equations, i.e.

$$\begin{aligned} \mathfrak{k}_E &= \ker \operatorname{ad}_E \cap \mathfrak{u}(N) \\ &= \{k \in \mathfrak{u}(N) \mid (\mathbb{1} \otimes E - E^\top \otimes \mathbb{1}) \operatorname{vec}(k) = 0\} \,. \end{aligned} \tag{199}$$

In particular, if E is of the form $E = \mu\mathbb{1} + \Omega$ with $\mu \in \mathbb{C}$ and $\Omega \in \mathfrak{u}(N)$, then $\mathfrak{k}_E$ is identical to the centraliser of Ω in $\mathfrak{u}(N)$.

By ortho-normalising the elements $k_j \in \mathfrak{k}_E$ of the generating set $\mathfrak{k}_E$ with $j = 1, 2, \ldots, n_E$, one obtains the projectors $P_j := |k_j\rangle\langle k_j|$ according to Eqn. (124) to give the total projection operator $P := \sum_j P_j$. With this definition, the gradient flow U2K of the summarising Table II applies and solves Eqn. (195). Therefore, the constraint of leaving a neutral state E invariant during the transfer from A to C can be approached *intrinsically* by restricting the flow from the full unitary group to a compact connected Lie subgroup, the stabiliser group $\mathbf{K}_E$ of E.

However, since it may be tedious to check for the stabiliser group $\mathbf{K}_E$ in each and every practical instance and then project the gradients onto the corresponding subalgebra $\mathfrak{k}_E$, a more versatile programming tool would be welcome. In [27], we therefore presented a combined approach based on the Lagrange-type function

$$L(U) = f_2(U) - \lambda\left(\operatorname{tr}\{E^\dagger UEU^\dagger\} - ||E||_2^2\right) \tag{200}$$

with $f_2(U) := |\operatorname{tr}\{C^\dagger UAU^\dagger\}|^2$. and *penalty term* $\lambda\left(\operatorname{tr}\{E^\dagger UEU^\dagger\} - ||E||_2^2\right)$. Here, the constraint $UEU^\dagger - E = 0$ was rewritten in the more convenient form $\operatorname{tr}\{E^\dagger UEU^\dagger\} - \|E\|_2^2 = 0$. The algorithm given in Table II as U2C implements a discretised gradient flow of L obtained from the identity

$$\begin{aligned} &D\,L(U)\,(\Omega U) \\ &\quad = \operatorname{tr}\left\{\left(2\left(f_2(U)^* \cdot [UAU^\dagger, C^\dagger]\right)_S - \lambda[UEU^\dagger, E^\dagger]\right)\Omega\right\}. \end{aligned}$$

Note that the *penalty parameter* λ is increased within the recursion to guarantee that the constraint is (at least approximately) satisfied in the limit.

Table II: Examples of Optimisation Tasks and Related Gradient Flows

no.	target function	discretised gradient flow	Ref.
I. Unconstrained Optimisation			
maximisation over the orthogonal group: $O \in SO(N,\mathbb{R})$ and $A, \Delta \in \mathbb{R}^{N\times N}$ with Δ diagonal, $\alpha_k > 0$ stepsize			
O1	$f(O) = \mathrm{tr}\{\Delta^\top O A O^\top\}$	$O_{k+1} = \exp\{-\alpha_k [O_k A O_k^\top, \Delta^\top]\} O_k$	[16, 21, 22]
maximisation over the unitary group: $U, V \in SU(N)$ and $A, C \in \mathbb{C}^{N\times N}$; $[\cdot,\cdot]_S$ and $(\cdot)_S$ denote skew-Hermitian parts			
U1	$f(U) = \mathrm{Re}\,\mathrm{tr}\{C^\dagger U A U^\dagger\}$	$U_{k+1} = \exp\{-\alpha_k [U_k A U_k^\dagger, C^\dagger]_S\} U_k$	[26, 27]
U2	$f(U) = \lvert\,\mathrm{tr}\{C^\dagger U A U^\dagger\}\rvert^2$	$U_{k+1} = \exp\{-\alpha_k \big([A_k, C^\dagger] f^*(U_k) - [A_k, C^\dagger]^\dagger f(U_k)\big)\} U_k$ where $A_k := U_k A U_k^\dagger$	[26, 27]
U3	$f(U,V) = \mathrm{Re}\,\mathrm{tr}\{C^\dagger U A V\}$	$U_{k+1} = \exp\{-\alpha_k (U_k A V_k C^\dagger)_S\} U_k$ $V_{k+1} = \exp\{-\beta_k (V_k C^\dagger U_k A)_S\} V_k$	[22, 28]
maximisation restricted to subgroups $\mathbf{K} \subset U(N)$ of the unitary group with $K \in \mathbf{K}$ and $P_\mathfrak{k}$ as projection from $\mathfrak{gl}(N,\mathbb{C})$ onto $\mathfrak{k}$, i.e. the Lie algebra to $\mathbf{K}$			
U1K	$f(K) = \mathrm{Re}\,\mathrm{tr}\{C^\dagger K A K^\dagger\}$	$K_{k+1} = \exp\{-\alpha_k P_\mathfrak{k}[K_k A K_k^\dagger, C^\dagger]\} K_k$	[here[a]]
U2K	$f(K) = \lvert\,\mathrm{tr}\{C^\dagger K A K^\dagger\}\rvert^2$	$K_{k+1} = \exp\{-\alpha_k \big(P_\mathfrak{k}[A_k, C^\dagger] f^*(K_k) - P_\mathfrak{k}[A_k, C^\dagger]^\dagger f(K_k)\big)\} K_k$ where $A_k := K_k A K_k^\dagger$	[here[a]]
U3K	$f(K_1, K_2) = \mathrm{Re}\,\mathrm{tr}\{C^\dagger K_1 A K_2\}$	$K_{k+1}^{(1)} = \exp\{-\alpha_k P_\mathfrak{k}(K_k^{(1)} A K_k^{(2)} C^\dagger)\} K_k^{(1)}$ $K_{k+1}^{(2)} = \exp\{-\beta_k P_\mathfrak{k}(K_k^{(2)} C^\dagger K_k^{(1)} A)\} K_k^{(2)}$	[28]
maximisation restricted to homogeneous spaces $\mathbf{G}/\mathbf{H}$ of the orthogonal group with $X \in \mathbf{G}/\mathbf{H}$ and A, C real symmetric			
O1P	$f(X) = \mathrm{tr}\{CX\}$ with $X_k := \mathrm{Ad}_{O_k}(A)$	$X_{k+1} = e^{-\alpha_k [X_k, C]}\, X_k\, e^{+\alpha_k [X_k, C]}$	[21, 115]
maximisation restricted to homogeneous spaces $\mathbf{G}/\mathbf{H}$ of the unitary group with $X \in \mathbf{G}/\mathbf{H}$ and A, C *arbitrary complex square* and $P_\mathfrak{k}$ as projection from $\mathfrak{gl}(N,\mathbb{C})$ onto $\mathfrak{k}$			
U1P	$f(X) = \mathrm{Re}\,\mathrm{tr}\{C^\dagger X\}$ with $X_k := \mathrm{Ad}_{U_k}(A)$	$X_{k+1} = e^{-\alpha_k [X_k, C^\dagger]_S}\, X_k\, e^{+\alpha_k [X_k, C^\dagger]_S}$	[here]
U1KP	$f(X) = \mathrm{Re}\,\mathrm{tr}\{C^\dagger X\}$ with $X_k := \mathrm{Ad}_{K_k}(A)$	$X_{k+1} = e^{-\alpha_k P_\mathfrak{k}[X_k, C^\dagger]}\, X_k\, e^{+\alpha_k P_\mathfrak{k}[X_k, C^\dagger]}$	[here]
II. Constrained Optimisation			
maximising $L(U)$ with penalty parameter $\lambda \in \mathbb{R}$ over the unitary group: $U \in SU(N)$; $A, C, D, E \in \mathbb{C}^{N\times N}$			
U1C	$L(U) = \mathrm{Re}\, f_C(U) - \lambda\, \mathrm{Im}^2 f_C(U)$ with $f_C(U) := \mathrm{tr}\{C^\dagger U A U^\dagger\}$	$U_{k+1} = \exp\{-\alpha_k \big([A_k, C^\dagger]_S + 2i\lambda\, \mathrm{Im}\, f_C(U_k) [A_k, C^\dagger]_H\big)\} U_k$ where $A_k := U_k A U_k^\dagger$ and $X_{H,S} := \frac{1}{2}(X \pm X^\dagger)$	[27]
U2C	$L(U) = \lvert f_C(U)\rvert^2 - \lambda(f_E(U) - \lvert\lvert E\rvert\rvert_2^2)$ $f_C(U)$ (s.a.) and $f_E(U) := \mathrm{tr}\{E^\dagger U E U^\dagger\}$	$U_{k+1} = \exp\{-\alpha_k \big((2 f_C^*(U_k)[A_k, C^\dagger])_S - \lambda [E_k, E^\dagger]\big)\} U_k$ where $A_k := U_k A U_k^\dagger$ and $E_k := U_k E U_k^\dagger$	[27]
U3C	$L(U) = \lvert f_C(U)\rvert^2 - \lambda \lvert f_D(U)\rvert^2$ $f_C(U)$ (s.a.) and $f_D(U) := \mathrm{tr}\{D^\dagger U A U^\dagger\}$	$U_{k+1} = \exp\{-2\alpha_k \big((f_C^*(U_k)[A_k, C^\dagger])_S - \lambda (f_D^*(U_k)[A_k, D^\dagger])_S\big)\} U_k$ where $A_k := U_k A U_k^\dagger$	[27]

[a] work presented in part at the MTNS 2006 [113]

Thus, for the constrained optimisation task of maximising the transfer from A to C subject to leaving the state E invariant, one has the choice of taking either the intrinsic approach U2K or the combined approach of U2C. Note, however, that the intrinsic approach restricts the flow to the stabiliser group $\mathbf{K}_E$ *at any time*, whereas the combined method is designed such as to start *arbitrarily* on $U(N)$ but *finally* to give an equilibrium point on $\mathbf{K}_E$. Therefore, the intrinsic approach has the advantage that the constraint is (at least in principal) properly satisfied for the entire iteration. However, there are situations where an intrinsic method is impractical as the computational costs are too expensive. The combined method, in contrast, does not suffer from this shortcom-

ing and thus has a wider range of applications. Note that the intrinsic approach paves the way to perform (or approximate) a transfer from A to C robustly by taking $\mathbf{K}_E$ as the stabiliser group resistent against a certain error class in the sense familiar from stabiliser codes [132, 133, 134]. The extrinsic approach, on the other hand, could be taken to transfer one protected state A to another one C via intermediate states that are no longer necessarily protected against errors as in the intrinsic case.

Finally, in [27, 109], we devised a penalty-type gradient flow algorithm for solving the constrained optimisation problem

$$\max_U |\operatorname{tr}\{C^\dagger U A U^\dagger\}| \quad \text{subject to} \quad \operatorname{tr}\{D^\dagger U A U^\dagger\} = \min .$$

To this end, we introduced the Lagrange-type function

$$L(U) := |\operatorname{tr}\{C^\dagger U A U^\dagger\}|^2 - \lambda\, |\operatorname{tr}\{D^\dagger U A U^\dagger\}|^2 \quad ,$$

to maximise the transfer from A to C while suppressing the transfer from A to D. This leads to the recursive scheme U3C in Table II.

V. CONCLUSIONS

The ability to calculate optima of quality functions for quantum dynamical processes and to determine steerings in concrete experimental settings that actually achieve these optima is tantamount to exploiting and manipulating quantum effects in future technology.

To this end, we have presented a comprehensive account of gradient flows on Riemannian manifolds (see general scheme of Fig. 9) allowing for generically convergent quantum optimisation algorithms—an ample array of explicit examples being given in Tab. II. Since the state space of quantum dynamical systems can often be represented by smooth manifolds, the unified foundations given are illustrated by many applications for numerically addressing two categories of problems: (a) we have focussed on abstract optimisation tasks over the dynamic group, and (b) we have sketched the relation to optimal control tasks in specified experimental settings.

In the present work on closed systems, a variety of applications are addressed by relating the dynamics to Lie group actions of the unitary group and its closed subgroups. Since symmetries give rise to stabiliser groups, particular emphasis has been on gradient flows on homogeneous spaces. Theory and algorithms have been structured and tailored for the following scenarios:

(i) for Lie groups with bi-invariant metric,

(ii) for closed subgroups

(iii) for compact Riemannian symmetric spaces,

or, more generally,

(iv) for naturally reductive homogeneous spaces.

As soon as the homogeneous spaces are no longer naturally reductive, the 'usual' way of obtaining geodesics on quotient spaces (by projecting geodesics from the group level to the quotient) fails. Alternatives of local approximations have been sketched in these cases in order to structure future developments.

Techniques based on the Riemannian exponential are easy to implement on Lie groups (with bi-invariant metric) and their closed subgroups. In particular, gradient flows on subgroups of the unitary group allow to address different partitionings of m-party quantum systems, the finest one being the group of purely local operations $SU(2)\otimes SU(2)\otimes\cdots\otimes SU(2)$. The corresponding gradient flows have several applications in quantum dynamics: for instance they prove useful tools to decide whether effective multi-qubit interaction Hamiltonians generate time evolutions that can be reversed in the sense of Hahn's spin echo solely by local operations.

As a new application, gradient flows on $SU(N_1) \otimes SU(N_2) \otimes \cdots \otimes SU(N_m)$ turned out to be a valuable and reliable alternative to conventional tensor-SVD methods for determining best rank-1 tensor approximations to higher-order tensors. In the case of m-party multipartite pure quantum states, they can readily be applied to optimising entanglement witnesses.

Double-bracket flows have been characterised as a special case of a broader class of gradient flows on naturally reductive homogeneous orbit spaces. Here, in view of using gradient techniques for ground-state calculations [36], it is important to note that double-bracket flows can also be established for any closed subgroup of $SU(N)$: by allowing for different partitionings $SU(N_1)\otimes SU(N_2)\otimes\cdots\otimes SU(N_m)$, one may set up a common frame to compare different types of unitary networks [36, 49] for calculating and simulating large-scale quantum systems.

Moreover, we have shown how techniques of restricting a gradient flow to subgroups also prove a useful tool for addressing constrained optimisation tasks by ensuring the constraints are fulfilled intrinsically. As an alternative, we have devised gradient flows that respect the constraints extrinsically, i.e., by way of penalty-type Lagrange parameters. These methods await application, e.g., in error-correction and robust state transfer.

Finally, in a follow-up study, we discuss the dynamics of open quantum systems in terms of Lie semigroups [58]. We sketch relations between the theory of Lie semigroups and completely positive semigroups. In particular in open systems, an easy characterisation of reachable sets arises only in very simple cases. It thus poses a current limit to an abstract optimisation approach on reachable sets. However, in these cases, gradient-assisted optimal control methods again prove valuable.

Therefore, not only does the current work give the justification to some recent developments, it also provides new techniques to the field of quantum control. It shows how to exploit the differential geometry in Lie theoretical terms for optimisation of dynamics on quantum-state manifolds. Thus the comprehensive theoretical treat-

ment illustrated by known examples and new practical applications has been given to fill a gap. — We anticipate the ample array of methods and their exemplifications will find broad application. The account of theoretical foundations is also meant to structure and trigger further basic research thus widening the set of useful tools.

VI. APPENDIX SECTION

Appendix A: Fréchet Differentials and Tangent Maps

For basic differential geometric terms and definitions we refer to [86]. Here, we recall only the fundamental relation between Fréchet derivatives and tangent maps for *finite dimensional* smooth manifolds.

Definition VI.1 (Fréchet derivative)**.** Let $\mathcal{H}$ and $\mathcal{H}'$ be real or comlpex Hilbert spaces. Let $F : \mathcal{U} \to \mathcal{H}'$ be a map defined in some open neighbourhood $\mathcal{U} \subset \mathcal{H}$ of X. Then F is *(Fréchet) differentiable* in X, if there exists a linear operator $DF(X) : \mathcal{H} \to \mathcal{H}'$ and a map $r_X : \mathcal{H} \to \mathcal{H}'$ such that

$$F(X + \Delta) - F(X) = DF(X)\Delta + r_X(\Delta)\|\Delta\| \qquad (201)$$

for $\Delta \in \mathcal{H}$ with $X + \Delta \in \mathcal{U}$ and $\|r_X(\Delta)\| \to 0$ for $\|\Delta\| \to 0$. The linear operator $DF(X)$ is called the *(Fréchet) derivative* of F in X and F is said to be *smooth*, if it is Fréchet differentiable of any order for all $X \in \mathcal{U}$.

Remark VI.1. (a) It is more common to stress the term 'Fréchet' derivative in an infinite dimensional setting to distinguish from other non-equivalent differentiability notions, cf. [135]. In finite dimensions, however, $DF(X)$ is often simply called 'the' derivative of F.

(b) Note, that if $DF(X)$ is complex linear for all $X \in \mathcal{U}$, then F is immediately holomorphic, i.e. F can be locally expanded in a power series, cf. [135].

Now, let M be a smooth manifold. If M is embedded in a Hilbert space $\mathcal{H}$, then the *tangent space* $T_X M$ to M at the point X is the set of all tangent vectors ξ at X, i.e. the set of all 'velocity' vectors

$$\xi = \frac{\mathrm{d}}{\mathrm{d}t}\gamma(t)\Big|_{t=0}$$

of smooth curves in M through $X = \gamma(0)$. If M is not embedded in a Hilbert space $\mathcal{H}$, the situation is slightly more complicated. Then, $T_X M$ is defined as the set of all equivalence classes $\xi = [\gamma]$ of smooth curves γ in M through $X = \gamma(0)$ having the same 'velocity' at $t = 0$, i.e. γ and $\widetilde{\gamma}$ represent the same equivalence class ξ if and only if

$$\frac{\mathrm{d}}{\mathrm{d}t}(\phi \circ \gamma)(t)\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}(\phi \circ \widetilde{\gamma})(t)\Big|_{t=0}$$

for one and thus for all charts ϕ at X. These equivalence classes ξ are called *tangent vectors*. Note that either way the tangent space $T_X M$ is a vector space isomorphic to the chart space.

Definition VI.2 (Tangent map)**.** Let M and N be smooth manifolds. Take $f : M \to N$ as a continuous map.

(a) Then f is called *smooth*, if $\psi \circ f \circ \phi^{-1}$ is smooth for all admissible charts ϕ and ψ.

(b) If f is smooth, the linear map $Df(x) : T_X M \to T_{f(X)}N$ given by

$$\xi = [\gamma] \mapsto Df(X)(\xi) := [f \circ \gamma]$$

is called the *tangent map* to $f : M \to N$ at $X \in M$.

Therefore, differentiability on manifolds is locally expressed in terms of charts. The associated tangent maps obey the standard rules of differential calculus, like the chain rule, etc.

Finally, we quote two useful results for computing tangent maps in the case of embedded manifolds.

Fact 1. If N is embedded in a Hilbert space $\mathcal{H}$, the above definition of the tangent map reads

$$\xi \mapsto Df(X)(\xi) = \frac{\mathrm{d}}{\mathrm{d}t}(f \circ \gamma)(t)\Big|_{t=0},$$

where γ denotes a representative of the tangent vector ξ.

Fact 2. Let $\mathcal{H}$, $\mathcal{H}'$ be Hilbert spaces and let $M \subset \mathcal{H}$, $N \subset \mathcal{H}'$ be embedded submanifolds. Furthermore, let $F : \mathcal{H} \to \mathcal{H}'$ be a smooth map with $F(M) \subset N$. Then the tangent map of $f := F|_M : M \to N$ at X is given by the restriction of the derivative $DF(X)$ to the tangent space of M at X, i.e.

$$Df(X) = DF(X)|_{T_X M} : T_X M \to T_{f(X)}N.$$

Appendix B: Polarisation Procedure

What is meant by the term *polarisation procedure*? Below, we give a short description of the concept and a sketch of its proof.

Let $\mathcal{H}$ be a real Hilbert space and $\beta : \mathbb{H} \to \mathbb{R}$ a bounded *quadratic form*, i.e., there exists a bounded symmetric bilinear form $B : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ such that

$$\beta(v) = B(v, v), \qquad (202)$$

for all $v \in \mathcal{H}$.

Claim A 1. *Let β be a bounded quadratic form. The bounded symmetric bi-linear form B which satisfies (202) is uniquely determined by β.*

Proof. By the symmetry and bi-linearity of B we have

$$B(v+w,v+w) = B(v,v) + B(w,w) + 2B(v,w)$$

and hence

$$\begin{aligned} B(v,w) &= \frac{1}{2}\Big(B(v+w,v+w) - B(v,v) - B(w,w)\Big) \\ &= \frac{1}{2}\Big(\beta(v+w) - \beta(v) - \beta(w)\Big) \end{aligned} \tag{203}$$

for all $v, w \in \mathcal{H}$. Therefore, B is uniquely determined by the quadratic form β. □

The above identity (203) is frequently called the *law of polarisation*. Next, we show that β can always be represented by a selfadjoint linear operator.

Claim A 2. *For any bounded quadratic form β there exists a unique bounded selfadjoint linear operator $\mathbf{B} : \mathcal{H} \to \mathcal{H}$ such that*

$$\langle v|\mathbf{B}v\rangle = \beta(v), \quad \textit{for all } v \in \mathcal{H}. \tag{204}$$

Proof. A straightforward application of the Riesz Representation Theorem yields that any bounded symmetric bi-linear form B on $\mathcal{H}$ can be represented by a bounded selfadjoint linear operator, cf. [136]. From this the result follows immediately. □

Remark A 1. If β is any form such that $\beta(tv) = t^2\beta(v)$ and $\beta(v) \leq K\|v\|^2$ is satisfied for all $t \in \mathbb{R}$ and $v \in \mathcal{H}$, then β is in general not induced by a bounded symmetric bi-linear form, or in other words, (203) does in general not define a bounded symmetric bi-linear form. However, if β meets the parallelogram identity, i.e.

$$\beta(v+w) + \beta(v-w) = 2\beta(v) + 2\beta(w) \tag{205}$$

for all $v, w \in \mathcal{H}$, then (203) does define a bounded symmetric bi-linear form.

Acknowledgments

Fruitful discussion with Jens Eisert on [36] is gratefully acknowledged. We wish to thank Otfried Gühne for drawing our attention to witness optimisation and Ref. [54]. This work was supported in part by the integrated EU programme QAP and by *Deutsche Forschungsgemeinschaft*, DFG, in the incentives SPP 1078 and SFB 631. Support and exchange enabled by the two Bavarian PhD programmes of excellence *Quantum Computing, Control, and Communication* (QCCC) as well as *Identification, Optimization and Control with Applications in Modern Technologies* is gratefully acknowledged.

[1] R. P. Feynman, Int. J. Theo. Phys. **21**, 467 (1982).
[2] R. P. Feynman, *Feynman Lectures on Computation* (Perseus Books, Reading, MA., 1996).
[3] A. Y. Kitaev, A. H. Shen, and M. N. Vyalyi, *Classical and Quantum Computation* (American Mathematical Society, Providence, 2002).
[4] P. W. Shor, in *Proceedings of the Symposium on the Foundations of Computer Science, 1994, Los Alamitos, California* (IEEE Computer Society Press, New York, 1994), pp. 124–134.
[5] P. W. Shor, SIAM J. Comput. **26**, 1484 (1997).
[6] R. Jozsa, Proc. R. Soc. A. **454**, 323 (1998).
[7] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, Proc. R. Soc. A. **454**, 339 (1998).
[8] M. Ettinger, P. Høyer, and E. Knill, Inf. Process. Lett. **91**, 43 (2004).
[9] L. K. Grover, in *Proceedings of the 28th Annual Symposium on the Theory of Computing, 1996, Philadelphia, Pennsylvania* (ACM Press, New York, 1996), pp. 212–219.
[10] L. K. Grover, Phys. Rev. Lett. **79**, 325 (1997).
[11] C. H. Papadimitriou, *Computational Complexity* (Addison Wesley, Reading, MA., 1995).
[12] S. Sachdev, *Quantum Phase Transitions* (Cambridge University Press, Cambridge, 1999).
[13] E. Jané, G. Vidal, W. Dür, P. Zoller, and J. Cirac, Quant. Inf. Computation **3**, 15 (2003).
[14] D. Porras and J. Cirac, Phys. Rev. Lett. **92**, 207901 (2004).
[15] J. Dowling and G. Milburn, Phil. Trans. R. Soc. Lond. A **361**, 1655 (2003).
[16] R. W. Brockett, in *Proc. IEEE Decision Control, 1988, Austin, Texas* (1988), pp. 779–803, see also: Lin. Alg. Appl., 146 (1991), 79–91.
[17] R. W. Brockett, Lin. Alg. Appl. **122-4**, 761 (1989).
[18] R. W. Brockett, Proc. Symp. Pure Math. **54**, 69 (1993).
[19] S. T. Smith, *Geometric Optimization Methods for Adaptive Filtering* (PhD Thesis, Harvard University, Cambridge MA, 1993).
[20] S. T. Smith, *Hamiltonian and Gradient Flows, Algorithms and Control* (American Mathematical Society, Providence, 1994), chap. Optimization Techniques on Riemannian Manifolds, pp. 113–136, Fields Institute Communications.
[21] U. Helmke and J. B. Moore, *Optimisation and Dynamical Systems* (Springer, Berlin, 1994).
[22] A. Bloch, ed., *Hamiltonian and Gradient Flows, Algorithms and Control*, Fields Institute Communications (American Mathematical Society, Providence, 1994).
[23] M. T. Chou and K. R. Driessel, SIAM J. Numer. Anal. **27**, 1050 (1990).
[24] P. A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds* (Princeton University Press, Princeton, 2008).
[25] L. Ambrosio and N. Gigli and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Lectures in Mathematics, ETH-Zürich (Birkhäuser, Basel, 2008), 2nd ed.

[26] S. J. Glaser, T. Schulte-Herbrüggen, M. Sieveking, O. Schedletzky, N. C. Nielsen, O. W. Sørensen, and C. Griesinger, Science **280**, 421 (1998).

[27] T. Schulte-Herbrüggen, *Aspects and Prospects of High-Resolution NMR* (PhD Thesis, Diss-ETH 12752, Zürich, 1998).

[28] T. Schulte-Herbrüggen and A. Spörl (2006), e-print: http://arXiv.org/pdf/quant-ph/0610061.

[29] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, J. Magn. Reson. **172**, 296 (2005).

[30] T. Schulte-Herbrüggen, A. K. Spörl, N. Khaneja, and S. J. Glaser, Phys. Rev. A **72**, 042331 (2005).

[31] A. K. Spörl, T. Schulte-Herbrüggen, S. J. Glaser, V. Bergholm, M. J. Storcz, J. Ferber, and F. K. Wilhelm, Phys. Rev. A **75**, 012302 (2007).

[32] T. Schulte-Herbrüggen, A. Spörl, N. Khaneja, and S. Glaser (2006), e-print: http://arXiv.org/pdf/quant-ph/0609037.

[33] P. Rebentrost, I. Serban, T. Schulte-Herbrüggen, and F. Wilhelm (2006), e-print: http://arXiv.org/pdf/quant-ph/0612165.

[34] M. Grace, C. Brif, H. Rabitz, I. Walmsley, R. Kosut, and D. Lidar, J. Phys. B.: At. Mol. Opt. Phys. **40**, S103 (2007).

[35] T. Schulte-Herbrüggen, A. K. Spörl, and S. J. Glaser (2007), e-print: http://arXiv.org/pdf/0712.3227.

[36] C. Dawson, J. Eisert, and T. J. Osborne, Phys. Rev. Lett. **100**, 130501 (2008), e-print: http://arXiv.org/pdf/0705.3456.

[37] M. Plenio, J. Eisert, J. Dreissig, and M. Cramer, Phys. Rev. Lett. **94**, 060503 (2005).

[38] M. Cramer, J. Eisert, M. Plenio, and J. Dreissig, Phys. Rev. A **73**, 012309 (2006).

[39] M. Wolf, F. Verstraete, M. B. Hastings, and I. Cirac (2007), e-print: http://arXiv.org/pdf/quant-ph/0704.3906.

[40] M. Fannes, B. Nachtergaele, and R. Werner, Lett. Math. Phys. **25**, 249 (1992).

[41] M. Fannes, B. Nachtergaele, and R. F. Werner, Commun. Math. Phys. **144**, 443 (1992).

[42] I. Peschel, X. Wang, M. Kaulke, and K. Hallberg, eds., *Density-Matrix Renormailzation: A New Numerical Method in Physics*, Lecture Notes in Physics Vol. 528 (Springer, Berlin, 1999).

[43] U. Schollwöck, Rev. Mod. Phys. **77**, 259 (2005).

[44] B. Schumacher and R. Werner (2004), e-print: http://arXiv.org/pdf/quant-ph/0405174.

[45] F. Verstraete, D. Porras, and I. Cirac (2004), e-print: http://arXiv.org/pdf/quant-ph/0404706.

[46] S. Anders, M. B. Plenio, W. Dür, F. Verstraete, and H. J. Briegel, Phys. Rev. Lett. **97**, 107206 (2006).

[47] G. Vidal, Phys. Rev. Lett. **99**, 220405 (2007).

[48] N. Schuch, M. Wolf, F. Verstraete, and I. Cirac (2007), e-print: http://arXiv.org/pdf/quant-ph/0708.1567.

[49] R. Hübner, C. Kruszynska, L. Hartmann, W. Dür, F. Verstraete, J. Eisert, and M. Plenio (2008), e-print: http://arXiv.org/pdf/0802.1211.

[50] F. Wegner, Ann. Phys. (Leipzig) **3**, 77 (1994).

[51] S. Kehrein, *The Flow-Equation Approach to Many-Particle Systems*, vol. 217 of *Springer Tracts in Physics* (Springer, Berlin, 2006).

[52] M. B. Plenio and S. Virmani, Quant. Comp. Inf. **7**, 1 (2007).

[53] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki (2007), e-print: http://arXiv.org/pdf/quant-ph/0702225.

[54] J. Eisert, P. Hyllus, O. Gühne, and M. Curty, Phys. Rev. A **70**, 062317 (2004).

[55] R. Lohmayer, A. Osterloh, J. Siewert, and A. Uhlmann, Phys. Rev. Lett. **97**, 260502 (2006).

[56] A. Osterloh, J. Siewert, and A. Uhlmann (2007), e-print: http://arXiv.org/pdf/0710.5909.

[57] C. Eltschka, A. Osterloh, J. Siewert, and A. Uhlmann (2007), e-print: http://arXiv.org/pdf/0711.4477.

[58] G. Dirr, U. Helmke, I. Kurniawan, and T. Schulte-Herbrüggen (2008), e-print: http://arXiv.org/pdf/0811.3906.

[59] M. M. Wolf and J. I. Cirac, Commun. Math. Phys. **279**, 147 (2008).

[60] C. Udrişte, *Convex Functions and Optimization Methods on Riemannian Manifolds* (Kluver, Dordrecht, 1994).

[61] D. Gabay, J. Optim. Theory Appl. **37(2)**, 177 (1982).

[62] M. Kleinsteuber, *Jacobi-Type Methods on Semisimple Lie Algebras—A Lie Algebraic Approach to the Symmetric Eigenvalue Problem* (PhD Thesis, Universität Würzburg, 2006).

[63] D. D'Alessandro, *Introduction to Quantum Control and Dynamics* (Chapman & Hall/CRC, Boca Raton, 2008).

[64] V. Jurdjevic, *Geometric Control Theory* (Cambridge University Press, Cambridge, 1997).

[65] A. E. Bryson and Y. C. Ho, *Applied Optimal Control: Optimization, Estimation, and Control* (Taylor and Francis, 1975).

[66] V. F. Krotov, *Global Methods in Optimal Control* (Marcel Dekker, New York, 1996).

[67] A. Peirce, M. Dahleh, and H. Rabitz, Phys. Rev. A **37**, 4950 (1987).

[68] K. L. Theo, C. J. Goh, and K. H. Wong, *A Unified Computational Approach to Optimal Control Problems* (Chapman & Hall/CRC, Boca Raton, 1991).

[69] Y. Maday and G. Turinici, J. Chem. Phys. **118**, 8191 (2003).

[70] G. Dirr and U. Helmke, GAMM-Mitteilungen **31**, 59 (2008).

[71] H. Sussmann and V. Jurdjevic, J. Diff. Equat. **12**, 95 (1972).

[72] V. Jurdjevic and H. Sussmann, J. Diff. Equat. **12**, 313 (1972).

[73] F. Colonius and W. Kliemann, *The Dynamics of Control* (Birkhäuser, Boston, 2000).

[74] R. W. Brockett, SIAM J. Control **10**, 265 (1972).

[75] R. W. Brockett, SIAM J. Appl. Math. **25**, 213 (1973).

[76] W. M. Boothby and E. N. Wilson, SIAM J. Control Optim. **17**, 212 (1979).

[77] F. Albertini and D. D'Alessandro, IEEE Trans. Automat. Control **48**, 1399 (2003).

[78] T. Schulte-Herbrüggen, K. Hüper, U. Helmke, and S. J. Glaser, *Applications of Geometric Algebra in Computer Science and Engineering* (Birkhäuser, Boston, 2002), chap. Geometry of Quantum Computing by Hamiltonian Dynamics of Spin Ensembles, pp. 271–283.

[79] M. W. Hirsch and S. Smale, *Differential Equations, Dynamical Systems, and Linear Algebra* (Academic Press, San Diego, 1974).

[80] M. C. Irwin, *Smooth Dynamical Systems* (Academic Press, New York, 1980).

[81] R. Fletcher, *Practical Methods of Optimization* (Wiley & Sons, Chichester, 1987), 2nd ed.

[82] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming*, Wiley-Interscience Series (Wiley & Sons, 1993), 2nd ed.

[83] W. Boothby, *An Introduction to Differential Manifolds and Riemannian Geometry* (Academic Press, New York, 1975).

[84] S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian Geometry*, Universitext (Springer, Berlin, 2004), 3rd ed.

[85] M. Spivak, *A Comprehensive Introduction to Differential Geometry, Vols. I-II* (Publish or Perish, Houston, 1999), 3rd ed.

[86] R. Abraham, J. E. Marsden, and T. Ratiu, *Manifolds, Tensor Analysis and Applications* (second edition, Springer, New York, 1988).

[87] J. Palis and W. de Melo, *Geometric Theory of Dynamical Systems* (Springer, New York, 1982).

[88] S. Łojasiewicz, *Sur les Trajectoires du Gradient d'une Fonction Analytique. Seminari di Geometria 1982-1983* (Università di Bologna, Istituto di Geometria, Dipartimento di Matematica, 1984).

[89] K. Kurdyka, Ann. Inst. Fourier **48**, 769 (1998).

[90] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry, Vols. I–II* (Wiley Interscience, New York, 1996).

[91] F. Takens, in *Manifolds - Amsterdam 1970*, edited by N. Kuiper (Springer, New York, 1971), Lecture Notes in Math., 197, p. 231.

[92] C. Lageman, *Convergence of Gradient-like Dynamical Systems and Optimization Algorithms* (PhD Thesis, Universität Würzburg, 2007).

[93] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces* (Academic Press, New York, 1978).

[94] A. Arvanitoyeorgos, *An Introduction to Lie Groups and the Geometry of Homogeneous Spaces* (American Mathematical Society, Providence, 2003).

[95] A. W. Knapp, *Lie Groups beyond an Introduction* (Birkhäuser, Boston, 2002), 2nd ed.

[96] B. C. Hall, *Lie Groups, Lie Algebras, and Representations* (Springer, New York, 2003).

[97] J. Milnor, Advances in Mathematics **21**, 293 (1976).

[98] J. J. Duistermaat and J. A. C. Kolk, *Lie Groups* (Springer, New York, 2000).

[99] J. Cheeger and D. G. Ebin, *Comparison Theorems in Riemannian Geometry* (North-Holland, Amsterdam, 1975).

[100] F. W. Warner, *Foundations of Differentiable Manifolds and Lie Groups* (Springer, New York, 1983).

[101] B. O'Neill, *Semi-Riemannian Geometry* (Academic Press, San Diego, 1983).

[102] T. Bröcker and T. tom Dieck, *Representation of Compact Lie Groups* (Springer, New York, 1985).

[103] O. Kowalski and J. Szenthe, Geom. Ded. **81**, 209 (2000), Erratum: ibid. **84**, 331 (2001).

[104] A. Besse, *Einstein Manifolds* (Spinger, Berlin, 1986).

[105] B. Kostant, Trans. Amer. Math. Soc. **80**, 520 (1955).

[106] U. Helmke, K. Hüper, and J. Trumpf (2007), e-print: http://arXiv.org/pdf/0709.2205.

[107] M. Goldberg and E. Straus, Lin. Alg. Appl. **18**, 1 (1977).

[108] C.-K. Li, Lin. Multilin. Alg. **37**, 51 (1994).

[109] T. Schulte-Herbrüggen, G. Dirr, U. Helmke, M. Kleinsteuber, and S. Glaser, Lin. Multin. Alg. **56**, 3 (2008).

[110] G. Dirr, U. Helmke, M. Kleinsteuber, and T. Schulte-Herbrüggen, Lin. Multin. Alg. **56**, 27 (2008).

[111] C.-K. Li and N. K. Tsing, Proc. Amer. Math. Soc. **111**, 19 (1991).

[112] U. Helmke, K. Hüper, J. B. Moore, and T. Schulte-Herbrüggen, J. Global Optim. **23**, 283 (2002).

[113] G. Dirr, U. Helmke, M. Kleinsteuber, S. Glaser, and T. Schulte-Herbrüggen, Proc. MTNS 2006, Kyoto (2006).

[114] G. Dirr, U. Helmke, M. Kleinsteuber, and T. Schulte-Herbrüggen, Proc. GAMM 2006, Berlin. PAMM **6**, 711 (2006).

[115] A. Bloch, R. W. Brockett, and T. Ratiu, Bull. Am. Math. Soc. **56**, 447 (1990).

[116] A. Bloch, R. W. Brockett, and T. Ratiu, Commun. Math. Phys. **147**, 57 (1992).

[117] R. Bertlman, H. Narnhofer, and W. Thirring, Phys. Rev. A **66**, 032319 (2002).

[118] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge (UK), 2000).

[119] O. Curtef, G. Dirr, and U. Helmke, Proc. ICIAM 2007, Zürich (submitted 2007).

[120] J. von Neumann, Tomsk Univ. Rev. **1**, 286 (1937), [reproduced in: *John von Neumann: Collected Works*, A.H. Taub, Ed., Vol. IV: Continuous Geometry and Other Topics, Pergamon Press, Oxford, 1962, pp 205-219].

[121] O. Sørensen, Prog. NMR Spectrosc. **21**, 503 (1989).

[122] J. Stoustrup, O. Schedletzky, S. J. Glaser, C. Griesinger, N. C. Nielsen, and O. W. Sørensen, Phys. Rev. Lett. **74**, 2921 (1995).

[123] D. Markham, J. A. Miszczak, Z. Puchała, and K. Życzkowski, Phys. Rev. A **77**, 042111 (2008).

[124] T. Kolda, SIAM J. Matrix Anal. Appl. **23**, 243 (2001).

[125] L. de Lathauwer, B. de Moor, and J. Vandewalle, SIAM J. Matrix Anal. Appl. **21**, 1253 (2000).

[126] L. de Lathauwer, B. de Moor, and J. Vandewalle, SIAM J. Matrix Anal. Appl. **21**, 1324 (2000).

[127] T. Wei and P. Goldbart, Phys. Rev. A **68**, 022307 (2003).

[128] T. G. Kolda and B. W.Bader (2005), Talk at Workshop on Tensor Decomposition and Applications, CIRM, Marseille, 2005.

[129] J. L. Brylinski, *Mathematics of Quantum Computation* (Chapman & Hall/CRC, Boca Raton, 2002), chap. Algebraic Measures of Entanglement, pp. 3–23, Computational Mathematics Series.

[130] B. G. Englert and N. Metwally, *Mathematics of Quantum Computation* (Chapman & Hall/CRC, Boca Raton, 2002), chap. Kinematics of Qubit Pairs, pp. 24–75, Computational Mathematics Series.

[131] F. Mezzadri, Notices Amer. Math. Soc. **54**, 592 (2007).

[132] M. Grassl, *Lectures on Quantum Information* (Wiley-VCH, Weinheim, 2007), chap. Quantum Error Correction, pp. 105–120.

[133] A. R. Calderbank, E. M. Rains, P. W. Shor, and N. J. A. Sloane, IEEE Trans. Inf. Theory **44**, 1369 (1998).

[134] A. R. Calderbank and P. W. Shor, Phys. Rev. A **54**, 1089 (1998).

[135] E. Hille and R. S. Phillips, *Functional Analysis and Semigroups* (American Mathematical Society, Providence, 1957).

[136] M. Reed and B. Simon, *Methods of Modern Mathe-*

matical Physics. Vol. I: Functional Analysis (Academic Press, San Diego, 1980).

[137] In general, Eqn. (41) is not used as a definition, but obtained as a consequence. Other definitions, e.g. via a linear connection, have the advantage that they guarantee bi-linearity straight away, cf. Remark III.1.

[138] Note that the coset-terminoloy in the group literature is not consistent, i.e. right cosets are sometimes called left cosets and vice versa. Here, we stick to the term right coset, if the group element in on the right side, i.e. $[G] = \mathbf{H}G$.

[139] A map φ is called *proper* if the pre-image $\varphi^{-1}(K)$ of any compact set K is also compact.