# Newton's Method for Path-Following Problems on Manifolds

Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrades der Bayerischen Julius-Maximilians-Universität Würzburg

vorgelegt von

Markus Baumann

aus

Erlenbach am Main

Würzburg 2008

Eingereicht am:

bei der Fakultät für Mathematik und Informatik
der Bayerischen Julius-Maximilians-Universität Würzburg

1. Gutachter: Prof. Dr. Uwe Helmke

2. Gutachter: Prof. Dr. Christian Kanzow

Tag der mündlichen Prüfung:

# Vorwort

Die vorliegende Arbeit wurde am Lehrstuhl 2 für Mathematik an der Universität Würzburg erstellt. Ein Forschungsschwerpunkt dieses Lehrstuhls ist die Entwicklung von Optimierungsalgorithmen auf Mannigfaltigkeiten. Passend dazu war der Ursprung dieser Arbeit, nämlich das Studium von Verfahren zur Bestimmung von orthogonalen Matrizen, welche vorgegebene symmetrische Matrizen diagonalisieren. Die besondere Schwierigkeit lag hierbei in der Tatsache, dass man sich nicht auf konstante Matrizen beschränkte, sondern symmetrische Matrizen betrachtete, die stetig differenzierbar von einem Parameter (der Zeit) abhingen. Für diese Problemstellung konnte man ein Verfahren zur zeitvarianten Berechnung ("Tracking") der gesuchten orthogonalen Transformation durch einen rein euklidischen Ansatz herleiten. Es stellte sich jedoch schnell heraus, dass der zugrunde liegende Trackingalgorithmus noch Verbesserungspotential besitzt. Zunächst sollte das Verfahren für Mannigfaltigkeiten verallgemeinert werden um strukturierte Trackingprobleme intrinsisch, d.h. direkt auf der Mannigfaltigkeit, behandeln zu können. Weitere Motive für Erweiterungen und Modifikationen der Trackingtheoreme waren die betrachteten Anwendungen. Dies waren zumeist Optimierungsprobleme auf Mannigfaltigkeiten, die bereits am Lehrstuhl behandelt wurden und welche man darüber hinaus in einem zeitvarianten Kontext untersuchen wollte.

Die Ausarbeitung dieser Ideen führte zu einigen Publikationen, die ich zusammen mit Herrn Prof. Dr. Uwe Helmke verfasste. An dieser Stelle möchte ich mich bei ihm ausdrücklich für seine Unterstützung bedanken. Ohne seine Erfahrung, sein Gespür für viel versprechende Untersuchungen, seine (mit-) anpackende und im positiven Sinne fordernde Art, wäre diese Arbeit nicht möglich gewesen. Weiteren Dank verdienen meine Kollegen und Freunde vom Lehrstuhl, bei denen man sich stets Unterstützung einholen konnte: Gunther Dirr, Jens Jordan, Martin Kleinsteuber und Christian Lageman sind hier in jedem Fall zu nennen. Besonderer Dank gilt auch Herrn Prof. Dr. Malte Messmann vom Juliusspital Würzburg, der mich in den ersten drei Promotionsjahren finanziell komfortabel ausgestattet hat. Leider konnten wir nicht wie ursprünglich geplant, ein interdisziplinäres medizinisch-mathematisches Thema zu einer Dissertation ausarbeiten. Trotzdem bekam ich genügend Freiheit, um dann ein rein mathematisches Thema zu behandeln. Und schließlich bedanke ich mich bei meinen Eltern, für ihre weit über das normale Maß hinausgehende Unterstützung während meiner Studienzeit.

# Publikationsliste

Die folgenden Arbeiten habe ich im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Mathematik II veröffentlicht.

- M. Baumann, U. Helmke. Diagonalization of time-varying symmetric matrices, *Proceedings of the International Conference on Computational Science ICCS*, 2002.
- M. Baumann, U. Helmke. Singular value decomposition of time-varying matrices, *Future Generation Computer Systems* 19(3):353-361, 2003.
- M. Baumann, U. Helmke. A path-following method for tracking of time-varying essential matrices, *Proceedings of the MTNS*, CD-Rom, Leuven, 2004.
- M. Baumann, C. Lageman, U. Helmke. Newton-type algorithms for time-varying pose estimation, *Proceedings of the ISSNIP*, CD-Rom, Melbourne, 2004.
- M. Baumann, J. Manton, U. Helmke. Reliable Tracking Algorithms for Principal and Minor Eigenvector Computations, *Proceedings of the 44th IEEE Conference on Decision and Control*, Sevilla, 2005.
- M. Baumann, U. Helmke. Riemannian subspace tracking algorithms on Grassmann manifolds, *Proceedings of the 46th IEEE Conference on Decision and Control*, New Orleans, 2007.
- M. Baumann, U. Helmke. A time-varying Newton algorithm for adaptive subspace tracking, *Mathematics and Computers in Simulation* , zur Veröffentlichung angenommen, 2008.

# Contents

# Chapter 1

# Introduction

## 1.1 Motivation

Many optimization problems for a smooth cost function $f : M \to \mathbb{R}$ on a manifold $M$ can be solved by determining the zeros $x_*$ of a vector field $F : M \to TM$ on $M$; such as e.g. the gradient $F$ of the cost function $f$. If $F$ does not depend on additional parameters, numerous zero-finding techniques are available for this purpose. It is a natural generalization however, to consider time-dependent optimization problems that require the computation of time-varying zeros $x_*(t)$ of time-dependent vector fields $F(x, t)$. Such parametric optimization problems arise in many fields of applied mathematics, in particular path-following problems in robotics [49], recursive eigenvalue and singular value estimation in signal processing, cf. [58], [59], [72], as well as numerical linear algebra and inverse eigenvalue problems in control theory cf. [53], [57] and [60]. In the literature, there are already some tracking algorithms for these tasks, but these do not always adequately respect the manifold structure. Hence, available tracking results can often be improved by implementing methods working directly on the manifold. For this reason, we develop in this thesis zero-finding techniques for time-varying vector fields on Riemannian manifolds $M$. Thus we consider a smooth parameterized vector field

$$F : M \times \mathbb{R} \to TM, \quad (x, t) \mapsto F(x, t) \in T_x M, \tag{1.1}$$

on a Riemannian manifold $M$. Hence $F$ can be regarded as a time-depending family of vector fields and the task is to determine a continuous curve $x_* : \mathbb{R} \to M$ such that

$$F(x_*(t), t) = 0 \tag{1.2}$$

holds for all $t \in \mathbb{R}$. This will be achieved by studying an extension of the Newton flow on manifolds. The discretization of the resulting ODE, which we will call time-varying Newton flow, then leads to concrete zero-tracking algorithms.

## 1.2 Previous work

Any numerical implementation of zero finding methods solving (1.2) requires discrete update schemes to compute estimates $x_k$ of the exact zero $x_*(t_k)$, where $t_k = kh$ for step size $h > 0$ and $k \in \mathbb{N}$. In Euclidean space, a simple choice for a zero tracking algorithm is obtained from the standard Newton method. The new estimate $x_{k+1}$ of $x_*(t_{k+1})$ is thus determined by a Newton update step for $F(x_k, t_{k+1})$, i.e. via

$$x_{k+1} = x_k - DF(x_k, t_{k+1})^{-1} F(x_k, t_{k+1}). \tag{1.3}$$

Thus, this method proceeds exactly as for time-invariant problems and therefore treat the time-dependency effects of the map only implicitly. This procedure obviously works as long as $x_k$ is in the domain of attraction of the zero of $F(\cdot, t_{k+1})$, which may not be the case for all $k \in \mathbb{N}$.

Continuation methods, cf. Allgower and Georg [4], [5], Garcia and Gould [26], Huitfeld and Ruhe [37], are natural tools for computing parameter-depending zeros of nonlinear maps. In the literature on continuation methods, the task of zero finding is mainly addressed in Euclidean space, i.e. for maps

$$F : \mathbb{R}^n \times [0, 1] \to \mathbb{R}^n, \quad (x, \lambda) \to F(x, \lambda),$$

where the task is to find $x_*(\lambda)$ such that $F(x_*(\lambda), \lambda) = 0$. This is achieved by studying the differential equation

$$DF(x, t) \cdot \dot{x} + \frac{\partial}{\partial t} F(x, t) = 0, \tag{1.4}$$

where $DF$ denotes the Fréchet derivative of $F$. Therefore, this approach may be suitable to solve the problem (1.2), but the main difference is, that we consider an unbounded interval for the second variable. Thus, known homotopy-type results can not be directly applied to our zero-finding task on $[0, \infty)$.

This is in contrast to the work of Davidenko [16], where the same differential equation (1.4) is considered on an unbounded time-interval for maps $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$. Under suitable full-rank conditions on $DF$, solutions $x(t)$ of this differential equation exist and are equal to the zero $x_*(t)$ of $F$, provided $x(0) = x_*(0)$. However, this algorithm only works under perfect initial conditions and does not perform any error correction. Thus, if one uses an Euler-step discretization of this algorithm to estimate the zero of $F$ at discrete times $t_k$, $k \in \mathbb{N}$, the integration error may accumulate at each step.

This leads us to the predictor/corrector methods, cf. [4], which calculate the zero $x_*(t)$ of $F$ at discrete times $t_k$ via two basic steps. The new iterate $x_{k+1}$ which approximates $x_*(t_{k+1})$ is obtained by

1. Compute a rough approximation $\tilde{x}_{k+1}$ of the zero of $F(\cdot, t_{k+1})$ using $x_k$ (predictor step).

2. Apply a zero-finding technique and the initial guess $\tilde{x}_{k+1}$ to get an improved estimate $x_{k+1} \approx x_*(t_{k+1})$ (corrector step).

A valid choice for the first step is e.g. the Euler-discretization of (1.4), which is given as
$$\tilde{x}_{k+1} = x_k - hDF(x_k, t_k)^{-1}\frac{\partial}{\partial t}F(x_k, t_k).$$
Hence, the evolution of $x_*(t)$ is linearly approximated to "predict" the new zero of $F$. The corrector step then consists of one or more iterates of the conventional Newton method:
$$x_{k+1} = \tilde{x}_{k+1} - DF(\tilde{x}_{k+1}, t_{k+1})^{-1}F(\tilde{x}_{k+1}, t_{k+1}).$$
The local quadratic convergence of Newton's algorithm guarantees that the resulting sequence has a reasonable accuracy, i.e. $x_k$ is a good approximation of the exact zero $x_*(t_k)$ for all $k \in \mathbb{N}$.
As it will turn out, our derived discrete tracking algorithms combine these two steps into one. This shows that our path-following method is closely related to such predictor/corrector algorithms. Moreover, our approach can be easily extended to a real predictor/corrector method, since the required additional Riemannian Newton step can be computed by a slight modification of our formulas.
Notably, our algorithms are derived for general problems on Riemannian manifolds and are therefore path-following methods on Riemannian manifolds. An other difference to the classical predictor/corrector methods is, that our tracking theorems are formulated such that they work with fixed step sizes instead of requiring intermediate step size adaptions.

The starting point of our work is the modified time-varying Newton flow
$$DF(x, t) \cdot \dot{x} + \frac{\partial}{\partial t}F(x, t) = -F(x, t), \tag{1.5}$$
which has been introduced in the PhD thesis of Getz [28], where the differential equation was derived by using the concept of the so-called "dynamic inverse". Solutions of this differential equation converge to the zero $x_*(t)$, where in particular, no perfect initial conditions are required. This is the benefit of inserting the additional feedback term into (1.4), which stabilizes the dynamics around $x_*(t)$. Hence, it is possible to introduce discrete versions of the above ODE such that the accuracy of the resulting sequence remains at a fixed level, cf. Section 2.3.1. An open question remained in how far one can extend Getz's method to a Riemannian manifold setting. This is exactly what will be studied in this thesis.

In order to develop Newton-type algorithms on manifolds, one can profit from recent publications about Riemannian optimization methods. For general information see e.g. Udistre [68] and Smith [63], where the latter studies Riemannian Newton methods, which are of particular interest for this thesis. Further results about Newton's algorithms on Riemannian manifolds can be found in Adler et al. [3], Dedieu and Priouret [17], Gabay [25], Ferreira and Svaiter [24] and in Mahony and Manton [46].
Since performing Newton-type methods on manifolds can be an expensive task, it is useful to provide ways to reconsider the problems in Euclidean spaces, which can sometimes be easier handled. The most popular way to do this, is to embed the

intrinsic problem into the ambient Euclidean space by using Lagrange multipliers, see e.g. Geiger and Kanzow [27] and Tanabe [65]. Note however, that such embeddings are not always easily available, and even if they are, Lagrange multiplier techniques may not work as defining equations for the manifolds are not always available.
An alternative to implement intrinsic Newton-type algorithms is to execute the zero-finding method on the tangent space and using parameterizations to get the corresponding updated point on the manifold, cf. Shub [61], Hüper and Trumpf [38] and Manton [47]. This idea will be discussed and extended further in this thesis, in order to obtain efficient implementations of the zero tracking algorithm. This allows even for the possibility of designing root finding techniques on arbitrary manifolds without referring to any Riemannian metric.

## 1.3 Results

As mentioned above, zero finding methods in the numerical literature have been mainly developed in an Euclidean space setting. Thus the methods start on a Riemannian submanifold $M \subset \mathbb{R}^n$ of Euclidean space and perform the computation steps in $\mathbb{R}^n$. In order to stay on a submanifold $M \subset \mathbb{R}^n$ such algorithms have to be combined with projection operations that map intermediate solutions in $\mathbb{R}^n$ onto the constraint set. This can be very cumbersome, technically involved and moreover, depends on an artificial choice of a suitable embedding of $M$. Thus, intrinsic methods are of interest that evolve during the entire computation on the manifold. It is the task of this thesis, to develop such intrinsic zero finding methods. The main results of this thesis are as follows:

- A new class of continuous and discrete tracking algorithms is proposed for computing zeros of time-varying vector fields on Riemannian manifolds.
- Convergence analysis is performed on arbitrary Riemannian manifolds.
- Concretization of these results on submanifolds, including for a new class of algorithms via local parameterizations.
- More specific results in Euclidean space are obtained by considering inexact and underdetermined time-varying Newton Flows.
- Illustration of these newly introduced algorithms by examining time-varying tracking tasks in three application areas.

The motivation of our studies is to provide continuous and discrete algorithms for tracking the smooth zero $x_*(t)$ of a time-varying vector field $F : M \times \mathbb{R} \to TM$. We introduce the **Riemannian time-varying Newton flow**

$$\nabla_{\dot{x}} F(x,t) + \frac{\partial}{\partial t} F(x,t) = \mathcal{M}(x) F(x,t), \tag{1.6}$$

defined by the covariant derivative $\nabla_{\dot{x}}F$ of $F$ along $x(t)$ with respect to the first variable. Here $\mathcal{M}$ denotes a stable bundle map, cf. Chapter 2 for details. Thus the setup is a generalization of equation (1.5) and we are able to extend and generalize the results and methods of Getz [28].

In Main Theorem 2.1, we derive sufficient conditions on $F$ such that the solution $x(t)$ of (1.6) asymptotically converges towards the zero $x_*(t)$ of $F$, i.e.

$$\operatorname{dist}(x(t), x_*(t)) \leq a\mathrm{e}^{-bt},$$

for some $a, b > 0$ and all $t \geq 0$. In particular, one does not need perfect initial conditions, since the zero of $F$ locally attracts solutions of the differential equation. This implies the local robustness of solutions of the dynamical system under perturbations.

The discretization of the time-varying Newton flow leads to an update scheme producing approximations $x_k$ for $x_*(t)$ at times $t_k = kh$ for step size $h > 0$ and $k \in \mathbb{N}$. Hence, the **time-varying Newton algorithm** is given for $\mathcal{M}(x) = -\frac{1}{h}I$ by

$$x_{k+1} = \exp_{x_k}\left((\nabla F(x_k, t_k))^{-1} \cdot \left(-F(x_k, t_k) - hF_\tau^h(x_k, t_k)\right)\right), \tag{1.7}$$

where $\exp_{x_k}$ is the exponential map of $M$ at $x_k \in M$, $(\nabla F(x_k, t_k))^{-1} : T_{x_k}M \to T_{x_k}M$ denotes the inverse of the covariant derivative of $F$ at $(x_k, t_k)$ and $F_\tau^h(x, t)$ is a step size-dependent approximation for $\frac{\partial}{\partial t}F(x, t)$, cf. Chapter 2 for details.

The second major result is formulated in Main Theorem 2.2, showing that the sequence $(x_k)_{k\in\mathbb{N}}$ of (1.7) has a guaranteed uniform accuracy holding in terms of the stepsize $h$. I.e. there exists a $c > 0$ such that for $h > 0$ holds that

$$\operatorname{dist}(x_k, x_*(t_k)) \leq ch,$$

for all $k \in \mathbb{N}$, provided that $\operatorname{dist}(x(0), x_*(0)) \leq ch$. This interesting feature is inherited from the local stability of the zero of $F$, together with a suitable choice of the feedback term $\mathcal{M}F$ in (1.6), and implies that the proposed algorithm has a convergence of order $h$ in $t$.

From these general results, more explicit versions of the continuous and discrete algorithms are derived for specific constraint sets: submanifolds, Lie groups and the Euclidean space. Even in the Euclidean case, it is useful to consider modifications of the Newton method, such as e.g. inexact or quasi-Newton methods, since a major difficulty of the tracking algorithms based on the time-varying Newton flow is the necessity of inverting the (covariant) derivative of $F$. If $F : \mathbb{R}^{n\times n} \times \mathbb{R} \to \mathbb{R}^{n\times n}$ is a matrix-valued map, one usually needs to invert a matrix representation of $DF(x, t)$, which has dimension $n^2 \times n^2$. To reduce computational effort, we consider inexact Newton flows, which are well studied in the time-invariant case, cf. [19]. We prove for time-varying maps $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$, that it suffices to approximatively invert the derivative of $F$ in order to achieve the tracking task. Thus we obtain a tracking algorithm

$$x_{k+1} = x_k - G(x_k, t_k)\left(F(x_k, t_k) + hF_\tau^h(x_k, t_k)\right),$$

where $G(x, t) : \mathbb{R}^n \to \mathbb{R}^n$ is a suitable approximation for the inverse $DF(x, t)^{-1} : \mathbb{R}^n \to \mathbb{R}^n$. Again, the resulting sequence has a guaranteed uniform accuracy $\leq ch$, cf.

Chapter 2 for details. Since such approximations can be computed with significantly less effort, this improves the applicability of the proposed algorithms. In particular, the need of computing the matrix representation of $DF$ may be dropped.
Finally, tracking results in Euclidean space are derived for underdetermined maps $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^m$, $m < n$. We call this the "underdetermined case", as the task becomes to track zero sets instead of single points. Analogously to the full rank map, we give conditions that the solutions of the underdetermined Newton flow asymptotically converge to the zero set of $F$, i.e. the distance of the solution of the time-varying Newton flow to the zero set decreases exponentially. Convergence properties are stated in Main Theorem 2.4, where also the inexact approach is included.
The algorithms defined in Euclidean space can also be used to solve intrinsic tracking tasks. The motivation for this is to circumvent the computation of the exponential map and geodesics, which may be quite complicated. Thus, it helps sometimes, to reformulate intrinsic tracking tasks in Euclidean spaces, either by including penalty terms for violating the constraints or by using Lagrange multipliers. Note however, that this requires a certain embedding of the original problem in an ambient Euclidean space, which causes other difficulties. In particular, the dimension of the occurring vectors in the algorithm usually increases.
Another way to implement intrinsic tracking algorithms may be used in the case where $F$ is the gradient of a cost function $\Phi : M \times \mathbb{R} \to \mathbb{R}$. Here we make use of suitable families of parameterizations $(\gamma_x)_{x\in M}$ and $(\mu_x)_{x\in M}$ of $M$ with $\gamma_x : V_x \to U_x \subset M$, $\mu_x : V_x \to U'_x \subset M$ and $\gamma_x(0) = \mu_x(0) = x$, where $V_x \subset \mathbb{R}^{\dim M}$. Then, the update scheme to track $x_*(t_k)$ is given by

$$x_{k+1} = \mu_{x_k}\left(-H_{\Phi\circ\hat{\gamma}_{x_k}}(0,t_k)^{-1}(\nabla(\Phi\circ\hat{\gamma}_{x_k})(0,t_k) + hG^h_{x_k}(0,t_k))\right), \tag{1.8}$$

where $\hat{\gamma}(y,t) := (\gamma(y),t)$, $\nabla(\Phi\circ\hat{\gamma}_{x_k})$ is the Euclidean gradient of $\Phi\circ\hat{\gamma}_{x_k}$ and $G^h_x(0,t)$ denotes an approximation of $\frac{\partial}{\partial t}\nabla(\Phi\circ\hat{\gamma}_x)(0,t)$, cf. Chapter 2 for details.
We prove that under certain conditions for $(\gamma_x)_{x\in M}$, $(\mu_x)_{x\in M}$ and $\Phi$, the accuracy of the resulting sequence satisfies $\operatorname{dist}(x_k, x_*(t_k)) \leq ch$ for some fixed $c > 0$ and all $k \in \mathbb{N}$, if $\operatorname{dist}(x_0, x_*(0)) \leq ch$, cf. Main Theorem 2.3. A special benefit of this approach is, that although it is an intrinsic algorithm, all computations are done with objects from Euclidean space and the dimension of these magnitudes does not exceed the dimension of the manifold.

The performance of these newly introduced methods is then evaluated by examining specific time-varying tracking problems.
In our first application in Chapter 3, we consider the task of **Intrinsic Subspace Tracking**, i.e. the computation of the principal (minor) subspace of a symmetric matrix, defined by the eigenspace corresponding to its largest (smallest) eigenvalues. Here, the natural state space for our algorithms is the Grassmann manifold $\mathrm{Grass}(m,n)$ of $m$-dimensional subspaces in $\mathbb{R}^n$.
For constant, time-invariant matrices, Edelman, Arias and Smith [23], Absil, Mahony and Sepulchre [2] proposed Riemannian optimization algorithms on Stiefel and Grassmann manifolds for principal subspace analysis. Their approach proceeds by applying

the Riemannian Newton method [63] to the Rayleigh quotient function. The same technique has been subsequently used by Lundström and Elden [44] for intrinsic subspace tracking of time-varying symmetric matrices. These approaches are equivalent to discretizations of the ordinary Newton flow, cf. (1.3). But due to the inherent lack of any tracking ability of this differential equation, no theoretical tracking bounds have been derived in [44].

In our approach, we go beyond that earlier work, by developing subspace tracking algorithms via the time-varying Newton algorithm (1.7) of the Rayleigh quotient function on a Grassmann manifold. Our algorithm achieves tracking with provable tracking bounds. By implementing particularly convenient parameterizations of the Grassmannian into the formulas, we obtain significantly simpler expressions of the discrete update scheme than given by [23, 44, 63]. Numerical experiments demonstrate the applicability and robustness of the proposed methods and a comparison with the methods of [44] is done.

In Chapter 4, we study the task of **Tracking Matrix Decompositions**. At first we consider the task of determining the eigenvalue and singular value decomposition of time-varying symmetric and non-square matrices, respectively. Eigenvalue decompositions of time-varying symmetric matrices $A(t) \in \mathbb{R}^{n\times n}$ have been studied by e.g. Dieci [20]. The authors derived matrix differential equations to track orthogonal transformations $X_*(t) \in O(n)$ such that $X_*(t)^\top A(t)X_*(t)$ is diagonal. However, these differential equations achieve asymptotic tracking of $X_*(t)$ only if they are exactly initialized, $X(0) = X_*(0)$. Moreover, concrete numerical implementations are missing, i.e. no discrete update scheme to compute the desired orthogonal matrices at discrete times is given. Since that approach bases on the homotopy method (1.4), a discrete version of the differential equation would require intermediate corrector steps, since discretization errors occur and accumulate at each step. In contrast to this, by using our tools of Chapter 2, robust update schemes are derived to perform the time-varying EVD of symmetic matrices in the cases of simple and multiple eigenvalues. Using a well known relation between the singular value and the symmetric eigenvalue problem, the developed diagonalizing method for symmetric matrices with multiple eigenvalues can be used to derive new SVD tracking algorithms of time-varying matrices $M(t) \in \mathbb{R}^{m\times n}$ for $m \geq n$, $t \in \mathbb{R}$. Notably, we used an approach basing on the inexact time-varying Newton flow to derive the EVD and SVD tracking methods. Therefore we did not need to vectorize the occurring matrices to obtain explicit update schemes, which considerably extends the use of these algorithms. Thus the maximal dimension of the matrices in the SVD tracking algorithm is $m \times m$, instead of $mn \times mn$. Numerical simulations at the end of the chapter confirm the theoretical robustness and good performance of the derived methods.

The polar decomposition of a full rank square matrix $M \in \mathbb{R}^{n\times n}$ is the factorization into an orthogonal and positive definite matrix. It is well-known, that a good method to compute the polar decomposition works via the SVD. Thus we propose to use the newly introduced SVD tracking algorithm for computing the polar factors. To assess the quality of this method, we compare it with an algorithm for the time-varying polar decomposition, which was introduced by Getz [28]. He derived a robust tracking

method basing on the time-varying Newton flow. Since the author did not give a concrete implementation, we applied our time-varying Newton algorithm to his setup. This leads to relatively large matrices in the algorithms, whose dimension is up to $\frac{1}{2}n(n+1) \times \frac{1}{2}n(n+1)$. Unlike these formulas, our SVD-basing algorithm works with matrices of the same dimension as $M$, i.e. $n \times n$. As expected, our method showed a better performance in the numerical examples and is therefore preferable.
In certain situations, it is not necessary to compute the whole eigenvalue decomposition of a symmetric matrix. Thus, there exists a number of methods, which determine only a few principal or minor eigenvectors of the matrix, cf. [55], [56], [58], [59]. In particular, a gradient based algorithm for the minor and principal eigenvector analysis was introduced in [48], which will be modified such that the time-varying Newton flow is applicable. Then we obtain algorithms tracking the time-varying minor and principal eigenvectors of time-varying symmetric matrices. Again, numerical results illustrate the good performance of the derived methods. Since principal eigenvector tracking algorithms can be employed to determine the principal eigenspace, we also compared these algorithms with the subspace tracking algorithms of Chapter 3. It turned out, that the algorithms of Chapter 3 have computationally advantages for general subspace tracking problems. This is due to the fact, that in contrast to the specific subspace tracking methods, the eigenvector tracking algorithms compute more information than required.
In Chapter 5, we consider an optimization problem, which arises in the area of computer vision: **Pose Estimation.** The task is to reconstruct the motion parameters $\Theta(t)$ (rotation) and $\Omega(t)$ (translation) of a rigid object, by evaluating time-varying image data. Our approach minimizes a suitable cost function on the manifold of so-called essential matrices, in analogy to [32] in the time-invariant case. Here, the extrinsic and intrinsic approaches as well as the parameterization method are applicable for the tracking task. We compare the different tracking methods and examine their specific difficulties and advantages while deriving the explicit update schemes. It turns out that the task of motion reconstruction can be achieved by all the derived algorithms, since all showed robust tracking results.

## 1.4 Notations

| | |
|---|---|
| $DF$ | Fréchet derivative of a map $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ |
| $M$ | Riemannian manifold |
| $T_xM$ | Tangent space of $M$ at $x \in M$ |
| $TM$ | Tangent bundle of $M$ |
| $\nabla_{\dot{x}} F(x(s), t)$ | Covariant derivative of a vector field $F : M \times \mathbb{R} \to TM$ at $(x(s), t) \in M \times \mathbb{R}$ along a curve $x : I \to M$ with respect to the first variable. |
| $\nabla F(x, t) : T_xM \to T_xM$ | Covariant derivative of a vector field $F : M \times \mathbb{R} \to TM$ at $(x, t) \in M \times \mathbb{R}$ with respect to the first variable |
| $\operatorname{grad} f$ | Intrinsic Riemannian gradient of a function $f : M \times \mathbb{R} \to \mathbb{R}$ with respect to the first variable |
| $H_f(x, t) \ : T_xM \to T_xM$ | Intrinsic Riemannian Hesse operator of $f : M \times \mathbb{R} \to \mathbb{R}$ with respect to the first variable |
| $B_r(0)$ | Neighborhood of 0 in the tangent space with radius $r$, i.e. $B_r(0) := \{v \in T_xM : \|v\| < r\}$ |
| $\mathcal{B}_r(x)$ | Intrinsic Riemannian neighborhood of $x \in M$ with radius $r$, i.e. $\mathcal{B}_r(x) := \{p \in M : \operatorname{dist}(x, p) < r\}$ |
| $\operatorname{dist}(x, y)$ | Intrinsic Riemannian distance of $x, y \in M$ |
| $\exp_x : T_xM \to M$ | Riemannian exponential map at $x \in M$ |
| $i_x(M)$ | Injectivity radius of the exponential map at $x \in M$ |
| $i^*(M)$ | Supremum of all $r \geq 0$ such that $\exp_x : B_r(0) \to M$ is a diffeomorphism for all $x \in M$ |
| $\operatorname{diag}(d_1, ..., d_n)$ | Diagonal matrix in $\mathbb{R}^{n\times n}$ with entries $d_1, ..., d_n$ |
| $\operatorname{rk}(A)$ | Rank of a matrix $A \in \mathbb{R}^{m\times n}$ |
| $\operatorname{tr}(A)$ | Trace of a matrix $A \in \mathbb{R}^{n\times n}$ |
| $A \otimes B$ | Kronecker product of matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$ |
| $\operatorname{VEC}(A)$ | Vectorizing operation of $A \in \mathbb{R}^{m\times n}$ by stacking the columns of $A$ under each other |
| $O(n)$ | Orthogonal group in $\mathbb{R}^{n\times n}$ |
| $(A)_Q$ | The $Q$-factor of the QR-factorization $A = (A)_Q R$ |
| Grass | Grassmann manifold |
| $\mathrm{Gr}_{m,n}$ | Isospectral representation of the Grassmann manifold (Grassmannian) |
| $\mathcal{E}_3$ | Normalized essential manifold |
| $\mathrm{Sym}_n$ | Vector space of all symmetric matrices of dimension $n \times n$ |
| $\mathfrak{so}_{\mathrm{n}}$ | Lie algebra of skew-symmetric matrices of size $n \times n$ |
| $[,]$ | Lie bracket, defined for matrices $A, B \in \mathbb{R}^{n\times n}$ by $[A, B] := AB - BA$ |
| $\operatorname{ad}_P \ : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ | Adjoint representation at $P$, i.e. $\operatorname{ad}_P(X) = [P, X]$ |

# Chapter 2

# Time-varying Newton flow

In this section we introduce the time-varying Newton flow, which is the mathematical basis of the zero tracking techniques considered in this work. At first, we study the dynamical system in the general case of working on a Riemannian manifold $(M, g)$. This leads to a continuous and a discrete version of the abstract tracking algorithm. Then we consider the special cases of $M$ being a Riemannian submanifold of $\mathbb{R}^n$ and a Lie group, leading to more concrete ODEs and update schemes. Thus, by using additional assumptions on $M$, we try to find better implementations of the considered tracking algorithms.
In the case of $M$ being a Riemannian submanifold, techniques to reformulate the tracking task in Euclidean space are studied. This extends the applicability of our approach, since one can use the Euclidean methods to solve the intrinsic tracking tasks then.
Finally, we also formulate the time-varying Newton flow in Euclidean space, leading to the most expressive versions of the algorithms. In order to improve the use of the studied methods, we show how to deal with inexact and underdetermined systems. Hence, practical modifications of the standard algorithm are derived, which are needed to implement the considered applications in the subsequent chapters.

## 2.1 Riemannian time-varying Newton flow

We now formulate the time-varying Newton flow on a Riemannian manifold. This requires an extension of standard tools from Riemannian geometry to an analysis of time-varying vector fields. Specifically, we discuss the Taylor formula for vector fields along curves on a Riemannian manifold. Although this can be done in a straightforward way, we summarize the required results in the next, preparatory section. For further references and details we refer to standard textbooks on Riemannian geometry, such as e.g. do Carmo [21].

### 2.1.1 Preliminaries on Riemannian manifolds

Let $M$ be a $k$–dimensional smooth manifold, endowed with a Riemannian metric $g$. A smooth vector field $X$ on $M$ then defines a smooth map from $M$ into the tangent bundle $TM$ that associates to each $p \in M$ a vector $X(p) \in T_pM$. We denote the set of all smooth vector fields on $M$ by $\mathcal{V}^\infty(M)$ and the set of all smooth functions on $M$ by $C^\infty(M)$. If $f : M \to N$ denotes a smooth map between manifolds, then for each point $p \in M$ the associated tangent map is denoted as $T_pf : T_pM \to T_{f(p)}N$. Let $\varphi : U \to V \subset \mathbb{R}^k$ denote a smooth coordinate chart on an open subset $U \subset M$. Given any smooth vector field $Y$ on $V$, the pull back vector field $\varphi^*Y : U \to TU$ is the smooth vector field on $U$ defined by

$$(\varphi^*Y)(p) = (T_p\varphi)^{-1}(Y(\varphi(p))),$$

where $T_p\varphi : T_pM \to T_{\varphi(p)}\mathbb{R}^k$ denotes the tangent map of $\varphi$. Let $\left\{\frac{\partial}{\partial y_1}, ..., \frac{\partial}{\partial y_k}\right\}$ denote the standard basis vectors of $\mathbb{R}^k$. Then a basis $\left\{\frac{\partial}{\partial \varphi_1}\Big|_p, ..., \frac{\partial}{\partial \varphi_k}\Big|_p\right\}$ of the tangent space $T_pM$, $p \in U$, is defined by

$$\frac{\partial}{\partial \varphi_i}\Big|_p := (T_p\varphi)^{-1}\frac{\partial}{\partial y_i}, \quad i = 1, ..., k. \tag{2.1}$$

Using this basis of $T_pM$ for $p \in M$, any vector field $X \in \mathcal{V}^\infty(M)$ is locally uniquely expressed as

$$X(p) = \sum_{i=1}^{k} a_i(p)\frac{\partial}{\partial \varphi_i}\Big|_p,$$

for $p \in U$ and smooth functions $a_i : U \to \mathbb{R}$, $i = 1, ..., k$.
The Lie derivative $L_X : C^\infty(M) \to C^\infty(M)$ of a vector field $X \in \mathcal{V}^\infty(M)$ is an $\mathbb{R}$–linear operator acting on smooth functions $f \in C^\infty(M)$ by $L_Xf := Xf$, defined by

$$(Xf)(p) = T_pf(X(p)), \quad p \in M.$$

Recall, that the Lie bracket product $[X, Y]$ of two vector fields $X, Y$ is the uniquely determined vector field satisfying

$$[X, Y]f = X(Yf) - Y(Xf)$$

for all $f \in C^\infty(M)$.

**Levi-Civita connection** In order to define derivatives of vector fields on $M$, we further need the concept of affine connections. We use the following notation. If $X, Y \in \mathcal{V}^\infty(M)$ are smooth vector fields on a Riemannian manifold $(M, g)$, then $\langle X, Y\rangle : M \to \mathbb{R}$ is the smooth function defined by

$$\langle X, Y\rangle(p) := g(p)(X(p), Y(p)), \quad p \in M.$$

**Definition 2.1.** *Let $f : M \to \mathbb{R}$ be a smooth function on a Riemannian manifold $(M, g)$ and let $df : M \to T^*M$ denote the associated differential one–form. The Levi-Civita connection $\nabla : \mathcal{V}^\infty(M) \times \mathcal{V}^\infty(M) \to \mathcal{V}^\infty(M)$ on $M$ assigns to each pair of smooth vector fields $X, Y$ on $M$ a smooth vector field $\nabla_XY$ with the properties:*

1. *$(X,Y) \mapsto \nabla_X Y$ is $\mathbb{R}$-bilinear in $X, Y$.*
2. *$\nabla_{fX} Y = f \nabla_X Y$ for all smooth functions $f$ on $M$.*
3. *$\nabla_X (fY) = f \nabla_X Y + (df X) Y$ for all smooth functions $f$ on $M$.*
4. *$\nabla_X Y - \nabla_Y X = [X, Y]$ (torsion freeness).*
5. *$d(\langle Y, Z \rangle) X = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle$ (compatibility with the metric).*

Note that a general affine connection does not require (4) and (5) of the previous definition. However, by including these additional assumptions, the Levi-Civita connection is uniquely defined and satisfies the Koszul formula.

$$2 \langle \nabla_X Y, Z \rangle = X \langle Y, Z \rangle + Y \langle X, Z \rangle - Z \langle X, Y \rangle - \langle Y, [X, Z] \rangle - \langle X, [Y, Z] \rangle + \langle Z, [X, Y] \rangle .$$

Any affine connection defines in a unique way the associated notion of a covariant derivative of differentiating a vector field along a smooth curve. Before turning to this crucial concept we recall that the affine connection $\nabla_X Y$ is asymmetrically defined with respect to $X, Y$. In fact, for any smooth vector fields $X, Y$ and $p \in M$, the value of $\nabla_X Y(p)$ depends only on the tangent vector $X(p)$ and the values of $Y$ on a neighborhood of $p$. Thus, for any tangent vector $v \in T_p M$ and any smooth vector field $X \in \mathcal{V}^\infty(M)$ with $X(p) = v$, the notation

$$\nabla_v Y(p) := \nabla_X Y(p)$$

is well defined and independent of the choice of such $X$. Let $c : I \to M$ denote a smooth curve. A smooth vector field $X(t)$ denote along $c$ is defined as a smooth map $X : I \to TM$, such that $X(t) \in T_{c(t)} M$ holds for all $t \in I$. We denote the infinite–dimensional vector space of all smooth vector fields along $c$ by $\mathcal{V}^\infty(c)$.

**Proposition 2.1.** *Let $M$ be a smooth manifold with an affine connection $\nabla$. Let $c : I \to M$ a smooth curve and $X, Y \in \mathcal{V}^\infty(c)$. Then there exists an unique $\mathbb{R}$–linear map $\frac{D}{dt} : \mathcal{V}^\infty(c) \to \mathcal{V}^\infty(c)$, called the covariant derivative along $c$, such that*

1. *$\frac{D}{dt}(X + Y) = \frac{DX}{dt} + \frac{DY}{dt}$.*
2. *$\frac{D}{dt}(fX) = \frac{df}{dt} X + f \frac{DX}{dt}$, where $f : I \to \mathbb{R}$ is a smooth function.*
3. *If $X$ is induced by a vector field $\hat{X} \in \mathcal{V}^\infty(M)$, i.e. $X(t) = \hat{X}(c(t))$ holds for all $t \in I$, then*
$$\frac{DX}{dt}(t) = \nabla_{\dot{c}(t)} \hat{X}(c(t)).$$
*Moreover, for all $n \in \mathbb{N}$ and $t \in I$*
$$\frac{D^n X}{dt^n}(t) = \nabla^n_{\dot{c}(t)} \hat{X}(c(t)).$$

Let $X, Y \in \mathcal{V}^\infty(c)$ be smooth vector fields along a smooth curve $c : I \to M$, expressed by using coordinate frames for some $\epsilon > 0$ and $t \in [0, \epsilon)$, i.e. $X(c(t)) = \sum_{i=1}^k X_i(c(t)) \frac{\partial}{\partial \varphi_i}\Big|_{c(t)}$ and $Y(c(t)) = \sum_{i=1}^k Y_i(c(t)) \frac{\partial}{\partial \varphi_i}\Big|_{c(t)}$. Then we get the local expression of the covariant derivative

$$\nabla_Y X(c(t)) = \sum_l \left( \sum_{i,j} X_i(c(t)) Y_j(c(t)) \Gamma_{ij}^l(c(t)) + dX_l(c(t)) Y(c(t)) \right) \frac{\partial}{\partial \varphi_l}\Big|_{c(t)}, \quad (2.2)$$

where $\Gamma_{ij}^l$ denote the Christoffel symbols of $\nabla$ for $i, j, l \in \{1, ..., k\}$.
Moreover, the covariant derivative of time-varying vector fields satisfies the following chain rule

**Lemma 2.1.** *Let $X : M \times \mathbb{R} \to TM$, $(p, t) \mapsto X(p, t)$, be a smooth time-varying vector field and let $c : I \to M$ a smooth curve. Thus $X(p, t) \in T_pM$ for $p \in M$ and all $t \in \mathbb{R}$. Then for $\tilde{X}(t) := X(c(t), t)$ and $X_t(p) := X(p, t)$*

$$\frac{D}{dt} \tilde{X}(t) = \nabla_{\dot{c}} X_t(c(t)) + \frac{\partial}{\partial t} X(c(t), t).$$

*Proof.* Let $\varphi : U \subset M \to V \subset \mathbb{R}^k$, $p \mapsto \varphi(p)$ a chart of $M$. Let $c : I \to U$ be a smooth curve. Then locally $X$ can be written as

$$X(p, t) = \sum_{i=1}^k a_i(p, t) \frac{\partial}{\partial \varphi_i}\Big|_p,$$

for some suitable smooth functions $a_i(p, t)$, $i = 1, ..., k$, cf. (2.1).
Thus for fixed $t$

$$\nabla_{\dot{c}} X_t(c(t)) = \frac{D}{ds} X(c(s), t) = \frac{D}{ds} \sum_{i=1}^k a_i(c(s), t) \frac{\partial}{\partial \varphi_i}\Big|_{c(s)} = \sum_{i=1}^k \frac{D}{ds} \left( a_i(c(s), t) \frac{\partial}{\partial \varphi_i}\Big|_{c(s)} \right) =$$

$$\sum_{i=1}^k \left( \frac{d}{ds} a_i(c(s), t) \right) \frac{\partial}{\partial \varphi_i}\Big|_{c(s)} + a_i(c(s), t) \frac{D}{ds} \frac{\partial}{\partial \varphi_i}\Big|_{c(s)} =$$

$$\sum_{i=1}^k da_i(c(s), t) \dot{c}(s) \frac{\partial}{\partial \varphi_i}\Big|_{c(s)} + a_i(c(s), t) \frac{D}{ds} \frac{\partial}{\partial \varphi_i}\Big|_{c(s)}$$

Therefore we get

$$\frac{D}{dt} \tilde{X}(t) = \frac{D}{dt} \sum_{i=1}^k a_i(c(t), t) \frac{\partial}{\partial \varphi_i}\Big|_{c(t)} = \sum_{i=1}^k \frac{D}{dt} \left( a_i(c(t), t) \frac{\partial}{\partial \varphi_i}\Big|_{c(t)} \right) =$$

$$\sum_{i=1}^k \left( \frac{d}{dt} a_i(c(t), t) \right) \frac{\partial}{\partial \varphi_i}\Big|_{c(t)} + a_i(c(t), t) \frac{D}{dt} \frac{\partial}{\partial \varphi_i}\Big|_{c(t)} =$$

$$\sum_{i=1}^{k} da_i(c(t),t)\dot{c}(t)\frac{\partial}{\partial\varphi_i}\Big|_{c(t)} + \frac{\partial}{\partial t}a_i(c(t),t)\frac{\partial}{\partial\varphi_i}\Big|_{c(t)} + a_i(c(t),t)\frac{D}{dt}\frac{\partial}{\partial\varphi_i}\Big|_{c(t)} =$$

$$= \frac{\partial}{\partial t}X(c(t),t) + \nabla_{\dot{c}}X_t(c(t)).$$

□

**Parallel translation** Let $Y$ be a smooth vector field along a smooth curve $c : [a,b] \to M$, $t \mapsto c(t)$. Then $Y$ is called parallel along $c(t)$ if

$$\frac{D}{dt}Y(t) = 0, \quad \forall t \in [a,b].$$

**Proposition 2.2.** *Let $M$ and $c$ be as above. Let $\xi \in T_{c(a)}M$ be a tangent vector of $M$ at $c(a)$. Then there exists a unique parallel vector field $V \in \mathcal{V}^\infty(c)$ such that $V(a) = \xi$.*

With the notation of above and $p = c(a)$, $q = c(s)$ for an arbitrary, but fixed $s \in [a,b]$, the parallel transport along $c$ induces a vector space isomorphism $\tau_{pq} : T_pM \to T_qM$, called the parallel translation, such that $\tau_{pq}\xi = V_\xi(s)$ for all $\xi \in T_pM$. Here $V_\xi$ is the parallel vector field along $c$ with $V_\xi(a) = \xi$.
Since the Levi-Civita connection is compatible with the metric, one has the following equation for vector fields $X, Y$ along curves $c : I \to M$:

$$\frac{d}{dt}\langle X, Y\rangle_{c(t)} = \langle \frac{DX}{dt}, Y\rangle_{c(t)} + \langle X, \frac{DY}{dt}\rangle_{c(t)}, \quad t \in I. \tag{2.3}$$

In particular for the Levi-Civita connection, one has $\langle X(t), Y(t)\rangle_{c(t)} = const$ for any smooth curve $c : I \to M$ and any pair of parallel vector fields $X, Y$ along $c$.

**Geodesics; complete manifold** Let $c : I \to M$ be a smooth, regular and injective curve. Here regularity means that $\dot{c}(t) \neq 0$ for all $t \in \mathbb{R}$. It is easily seen that there exists then a smooth vector field $X : M \to TM$ such that $\dot{c}(t) = X(c(t))$ for all $t \in I$. Thus,

$$\nabla_{\dot{c}}\dot{c} := (\nabla_X X)(c(t))$$

and $c$ is a geodesic, if $\nabla_{\dot{c}}\dot{c} = 0$ for all $t \in I$. Then $\|\dot{c}(t)\|_{c(t)} = \|\dot{c}(0)\|_p$ for all $0 \leq t \leq 1$, which can be easily seen by using (2.3).
We always require that $M$ is a geodesically complete manifold, i.e. for all $p \in M$, any geodesic $c(t)$ starting from $p$ is defined for all $t \in \mathbb{R}$. Hence, the exponential map at $p \in M$ is defined for any $v \in T_pM$ by

$$\exp_p(v) = c(1),$$

where $c : \mathbb{R} \to M$ is a geodesic with $c(0) = p$ and $\dot{c}(0) = v$. Note that every compact Riemannian manifold is geodesically complete. The injectivity radius $i_p(M)$ at a point $p \in M$ is the supremum of $r > 0$ such that $\exp_p\big|_{B_r(0)}$ is injective. Here, $B_r(0) = \{v \in T_pM \mid \|v\|_p < r\}$. We further use the notation $i_p^*(M)$ at a point $p \in M$, which denotes the supremum of $r > 0$ such that $\exp_p \big|_{B_r(0)}$ is a diffeomorphism. Since $i_p(M)$ can vary with the base point $p$, we need to define as well the global injectivity radius $i(M)$ of $M$ by

$$i(M) := \inf\{i_p(M)|\ p \in M\},$$

as well as

$$i^*(M) := \inf\{i^*_p(M)|\ p \in M\}$$

for the largest radius of the ball, where $\exp_p$ is a local diffeomorphism for all $p \in M$. Note that for a compact Riemannian manifold $M$, the global numbers $i(M)$ and $i^*(M)$ are always positive.
We denote the intrinsic distance of $p, q \in M$ by

$$\operatorname{dist}(p,q) := \inf\left\{\int_0^1 \|\dot{c}(t)\|_{c(t)} dt \middle|\ c\colon \mathbb{R} \to M \ \text{ piecewise } C^1 \text{ curve, } c(0) = p,\ c(1) = q\right\}$$

Let $p, q \in M$ and let $c : \mathbb{R} \to M$ be a geodesic form $p$ to $q$ with $c(0) = p$ and $c(1) = q$. If $\|\dot{c}(0)\|_p \leq i^*_p(M)$, then $\exp_p^{-1}(q)$ is defined with

$$\exp_p^{-1}(q) = \dot{c}(0).$$

Moreover, $c$ is the unique length minimizing geodesic from $p$ to $q$ and

$$\operatorname{dist}(p,q) = \|\dot{c}(0)\|_p.$$

We define the intrinsic neighborhood of $p \in M$ by $\mathcal{B}_R(p) := \{q|\ \operatorname{dist}(p,q) < R\}$ for $R > 0$. If $M$ is geodesically complete and $0 < R \leq i^*_p(M)$, then $\mathcal{B}_R(p) = \{\exp_p v|\ v \in T_pM, \|v\|_p < R\}$.

**Riemannian Hesse operator** Let $f : M \to \mathbb{R}$ be a smooth function on a Riemannian manifold $(M, g)$. The gradient vector field of $f$ is the uniquely defined smooth vector field $\operatorname{grad} f$ on $M$, characterized by

$$df(p) \cdot \xi = \langle \operatorname{grad} f(p), \xi \rangle_p\,,$$

for all $\xi \in T_pM$ and $p \in M$. On a Riemannian manifold one also has the intrinsic analogue of the second derivative, the Hessian. The Hesse operator of $f$ is defined as follows, cf. [34].

**Definition 2.2.** *a) The Hesse form of a smooth function $f : M \to \mathbb{R}$ at a critical point $p \in M$ is the symmetric bilinear form on the tangent space*

$$\mathcal{H}_f(p) : T_pM \times T_pM \to \mathbb{R},$$

$$(\xi,\eta) \mapsto \frac{1}{2}\big(\mathcal{H}_f(p)(\xi+\eta,\xi+\eta) - \mathcal{H}_f(p)(\xi,\xi) - \mathcal{H}_f(p)(\eta,\eta)\big)$$

*defined for tangent vectors $\xi, \eta \in T_pM$ via the quadratic form*

$$\mathcal{H}_f(p)(\xi,\xi) := (f \circ \alpha)''(0).$$

*Here $\alpha : I \to M$ is an arbitrary curve with $\alpha(0) = p$ and $\dot{\alpha}(0) = \xi$.*

b) *The Riemannian Hesse form of a smooth function $f : M \to \mathbb{R}$ at an arbitrary point $p \in M$ is the symmetric bilinear form on the tangent space*

$$\mathcal{H}_f(p) : T_pM \times T_pM \to \mathbb{R},$$

$$(\xi, \eta) \mapsto \frac{1}{2}\big(\mathcal{H}_f(p)(\xi+\eta, \xi+\eta) - \mathcal{H}_f(p)(\xi, \xi) - \mathcal{H}_f(p)(\eta, \eta)\big)$$

*defined for tangent vectors $\xi, \eta \in T_pM$ via the quadratic form*

$$\mathcal{H}_f(p)(\xi, \xi) := (f \circ c)''(0).$$

*Here $c : I \to M$ denotes the (locally) unique geodesic, with $c(0) = p$ and $\dot{c}(0) = \xi$. The Riemannian Hesse operator then is the uniquely determined selfadjoint map*

$$H_f(p) : T_pM \to T_pM$$

*satisfying*

$$\mathcal{H}_f(p)(\xi, \eta) = \langle H_f(p)\xi, \eta\rangle_p$$

*for all tangent vectors $\xi, \eta \in T_pM$.*

**Remark 2.1.** Note that the first part of the above definition applies only at critical points. To define the Hessian at an arbitrary point, we need the additional structure of a Riemannian metric to define geodesics, as done in the second part.
Note further that the above two definitions of the Hesse form coincide at a critical point. In particular, the Riemannian Hesse form at a critical point is independent of the choice of the Riemannian metric.

Instead of defining the Hessian in this way by using geodesics, one can also give an equivalent definition in terms of the Levi-Civita connection.

**Proposition 2.3.** *Let $p \in M$, $\xi, \eta \in T_pM$ be tangent vectors and $X, Y$ be smooth vectors fields on $M$ with $X(p) = \xi, Y(p) = \eta$. Then the Riemannian Hesse form and Hesse operator, respectively, are given as*

$$\mathcal{H}_f(p)(\xi, \eta) = X(Yf)(p) - ((\nabla_X Y)f)(p) \tag{2.4}$$

*and*

$$H_f(p)\xi = (\nabla_X \mathrm{grad}\, f)(p). \tag{2.5}$$

*In particularly, for any smooth curve $c : I \to M$*

$$\nabla_{\dot{c}}\mathrm{grad}\, f(c(t)) = H_f(c(t))\dot{c}(t).$$

**Taylor's formula on Riemannian manifolds** In the sequel we will need to estimate the difference of a vector field $X$ at two different points $p, q \in M$. In order to do so we use Taylor expansions on a manifold. The proof of the following characterization of the covariant derivative in terms of parallel transport can be found in the book of Helgason [31].

**Theorem 2.1.** *Let $M$ be a manifold with an affine connection $\nabla$. Let $p \in M$ and let $X, Y \in \mathcal{V}^\infty(M)$. Let $t \mapsto c(t)$ be an integral curve of $X$ through $p = c(0)$ and $\tau_{pc(t)}$ the parallel translation from $p$ to $c(t)$ with respect to the curve $c$. Then*

$$\nabla_X Y(p) = \lim_{t \to 0} \frac{1}{t}(\tau_{pc(t)}^{-1} Y(c(t)) - Y(p)).$$

The following lemma uses this result to derive formulas for higher order covariant derivatives.

**Lemma 2.2.** *Let $c : [a, b] \to M$ be a smooth regular curve, i.e. $\dot{c}(t) \neq 0$ for all $t \in [a, b]$. Let further $X \in \mathcal{V}^\infty(c)$. Then*

$$\tau_{pc(t)}^{-1} \frac{D^i}{dt^i} X(t) = \frac{d^i}{ds^i} \tau_{pc(t+s)}^{-1} X(t+s)\Big|_{s=0}$$

*where $p = c(a)$.*

*Proof.* Let $q = c(t)$, $r = c(t+s)$. We prove the claim inductively for $i$: For $i = 1$, we have

$$\frac{d}{ds}\tau_{pr}^{-1} X(t+s)\Big|_{s=0} = \frac{d}{ds}\tau_{pq}^{-1}\tau_{qr}^{-1} X(t+s)\Big|_{s=0} = \tau_{pq}^{-1}\frac{d}{ds}\tau_{qr}^{-1} X(t+s)\Big|_{s=0} = \tau_{pq}^{-1}\frac{D}{dt} X(t).$$

Note that the last equality follows from the previous theorem.
Thus let the claim be true for some $i \in \mathbb{N}$. Then

$$\frac{d^{i+1}}{ds^{i+1}}\tau_{pr}^{-1} X(t+s)\Big|_{s=0} = \frac{d^{i+1}}{ds^{i+1}}\tau_{pq}^{-1}\tau_{qr}^{-1} X(t+s)\Big|_{s=0} = \tau_{pq}^{-1}\frac{d^{i+1}}{ds^{i+1}}\tau_{qr}^{-1} X(t+s)\Big|_{s=0} =$$
$$\tau_{pq}^{-1}\frac{d}{ds}\left(\frac{d^i}{ds^i}\tau_{qr}^{-1} X(t+s)\right)\Big|_{s=0} = \tau_{pq}^{-1}\frac{D}{dt}\frac{D^i}{dt^i} X(t) = \tau_{pq}^{-1}\frac{D^{i+1}}{dt^{i+1}} X(t),$$

which completes the proof. □

Using this lemma, we derive Taylor's formula for vector fields.

**Theorem 2.2.** *Let $X \in \mathcal{V}^\infty(M)$ and let $c : [0, r] \to M$ a smooth regular curve from $p = c(0)$ to $q = c(r)$. Let $\tilde{X}(t) := X(c(t))$. Then for all $t \in [0, r]$ and $n \in \mathbb{N}$*

$$\tau_{pc(t)}^{-1}\tilde{X}(t) = \sum_{i=0}^{n-1}\frac{t^i}{i!}\frac{D^i}{dt^i}\tilde{X}(0) + \frac{t^n}{(n-1)!}\int_0^1 (1-s)^{n-1}\tau_{pc(ts)}^{-1}\frac{D^n}{du^n}\tilde{X}(u)\Big|_{u=ts} ds.$$

*In particular, we obtain Taylor's formula for smooth vector fields*

$$\tau_{pq}^{-1} X(q) = \sum_{i=0}^{n-1}\frac{r^i}{i!}\nabla_{\dot{c}(0)}^i X(p) + \frac{r^n}{(n-1)!}\int_0^1 (1-s)^{n-1}\tau_{pc(rs)}^{-1}\nabla_{\dot{c}(rs)}^n X(c(rs)) ds. \quad (2.6)$$

*Proof.* Let $Y : [0, r] \to V$ be a smooth curve in a finite dimensional $\mathbb{R}$-vector space $V$. The Taylor Theorem implies

$$Y(t) = Y(0)+t\frac{d}{dt}Y(0)+...+\frac{t^{n-1}}{(n-1)!}\frac{d^{n-1}}{dt^{n-1}}Y(0)+\frac{t^n}{(n-1)!}\int_0^1(1-s)^{n-1}\frac{d^n}{du^n}Y(u)\Big|_{u=ts}ds$$

Hence for $Y(t) := \tau^{-1}_{pc(t)}\tilde{X}(t) \in T_pM$ and $p = c(0)$ we obtain from Lemma 2.2

$$\frac{d^i}{ds^i}Y(0) = \tau^{-1}_{pc(t)}\frac{D^i}{dt^i}\tilde{X}(t)$$

and therefore

$$\tau^{-1}_{pc(t)}\tilde{X}(t) =$$

$$\tilde{X}(0)+t\frac{D}{dt}\tilde{X}(0)+...+\frac{t^{n-1}}{(n-1)!}\frac{D^{n-1}}{dt^{n-1}}\tilde{X}(0)+\frac{t^n}{(n-1)!}\int_0^1(1-s)^{n-1}\tau^{-1}_{pc(ts)}\frac{D^n}{du^n}\tilde{X}(u)\Big|_{u=ts}ds.$$

Thus we get for $t = r$

$$\tau^{-1}_{pq}\tilde{X}(r) =$$

$$\tilde{X}(0) + r\frac{D}{dt}\tilde{X}(0) + \frac{r^{n-1}}{(n-1)!}\frac{D^{n-1}}{dt^{n-1}}\tilde{X}(0) + \frac{r^n}{(n-1)!}\int_0^1(1-s)^{n-1}\tau^{-1}_{pc(rs)}\frac{D^n}{du^n}\tilde{X}(u)\Big|_{u=rs}ds.$$

The result follows from claim 3) of Proposition 2.1. □

**Remark 2.2.** (Estimate for 2 variables)
By using Taylor's formula once more, we obtain an approximation formula for vector fields, which depends on an additional variable $t$.
Thus let $X : M\times\mathbb{R} \to TM$ be a time-varying vector field, defined by $(p, t) \mapsto X(p, t) \in T_pM$, and consider for $q \in M$

$$X(q, t_1) = X(q, t_0) + \frac{\partial}{\partial t}X(q, t)\big|_{t=t_0}h + \int_0^1(1 - s)\frac{\partial^2}{\partial t^2}X(q, t)\big|_{t=t_0+sh}h^2ds,$$

where $h = t_1 - t_0$.
Let $c : [0, r] \to M$ a smooth regular curve from $p = c(0)$ to $q = c(r)$. Equation (2.6) reads for $n = 2$ and $X = X(p, t)$

$$\tau^{-1}_{pq}X(q, t_0) = X(p, t_0) + r\nabla_{\dot{c}(0)}X(p, t_0) + r^2\int_0^1(1 - s)\tau^{-1}_{pc(rs)}\nabla^2_{\dot{c}(rs)}X(c(rs), t_0)ds.$$

Hence

$$\tau^{-1}_{pq}X(q, t_1) = X(p, t_0) + r\nabla_{\dot{c}(0)}X(p, t_0) + r^2\int_0^1(1 - s)\tau^{-1}_{pc(rs)}\nabla^2_{\dot{c}(rs)}X(c(rs), t_0)ds+$$

$$+\frac{\partial}{\partial t}\left(X(p, t) + r\int_0^1\tau^{-1}_{pc(rs)}\nabla_{\dot{c}(rs)}X(c(rs), t)ds\right)\Big|_{t=t_0}h+$$

$$\tau_{pq}^{-1}\int_0^1(1-s)\frac{\partial^2}{\partial t^2}X(q,t)\big|_{t=t_0+sh}h^2ds,$$

where we used the Taylor formula for $n=1$, i.e.

$$\tau_{pq}^{-1}X(q,t)=X(p,t)+r\int_0^1\tau_{pc(rs)}^{-1}\nabla_{\dot{c}(rs)}X(c(rs),t)ds.$$

Therefore, we get

$$\boxed{\tau_{pq}^{-1}X(q,t_1)=X(p,t_0)+r\nabla_{\dot{c}(0)}X(p,t_0)+h\tfrac{\partial}{\partial t}X(p,t_0)+\mathcal{R}} \tag{2.7}$$

where the remainder term is

$$\mathcal{R}=r^2\int_0^1(1-s)\tau_{pc(rs)}^{-1}\nabla_{\dot{c}(rs)}^2X(c(rs),t_0)ds+rh\frac{\partial}{\partial t}\int_0^1\tau_{pc(rs)}^{-1}\nabla_{\dot{c}(rs)}X(c(rs),t)ds\Big|_{t=t_0}+$$

$$h^2\tau_{pq}^{-1}\int_0^1(1-s)\frac{\partial^2}{\partial t^2}X(q,t)\big|_{t=t_0+sh}ds$$

**Inspecting the second order covariant derivative**

In order to apply Theorem 2.2 to get estimates for the covariant derivatives, we need some further examinations. Thus let $p\in M$, $R\leq i_p^*(M)$ and $\varphi:=\left(\exp_p|_{B_R(0)}\right)^{-1}$ a chart of $M$. Let $q\in\mathcal{B}_R(p)$ and $v:=\exp_p^{-1}q$. Thus, $\|v\|_p=\operatorname{dist}(p,q)$ and $c(t):=\exp_p(tv)$ is a geodesic joining $p$ and $q$ for $t\in[0,1]$.
For any fixed $w\in T_qM$ we define a smooth vector field $Y_w$ on $\mathcal{B}_R(p)$ by

$$Y_w(p'):=\tau_{pp'}\tau_{pq}^{-1}w,$$

where $\tau_{pp'}$ denotes the parallel transport along the unique geodesic joining $p$ and $p'\in\mathcal{B}_R(p)$. Thus, along geodesics $c:[0,1]\to M$ with $c(0)=p$, we have $Y_w(c(t))=\tau_{pc(t)}\tau_{pq}^{-1}w$, which is a parallel vector field along $c$, i.e. $\frac{D}{dt}Y_w(c(t))=0$.
To get an approximation formula for $\nabla_{\dot{c}}\nabla_{Y_w}X(c(t))$ we locally express the occurring vector fields by the coordinate frames associated with $\varphi$. Hence consider the smooth function $X_i(p')$, $Y_i(p')$ and $C_i(p')$ $(i=1,...,k)$ on $\mathcal{B}_R(p)$ such that $X(p')=\sum_{i=1}^kX_i(p')\frac{\partial}{\partial\varphi_i}\Big|_{p'}$, $Y_w(p')=\sum_{i=1}^kY_i(p')\frac{\partial}{\partial\varphi_i}\Big|_{p'}$ and $\dot{c}(t)=\sum_{i=1}^kC_i(c(t))\frac{\partial}{\partial\varphi_i}\Big|_{c(t)}$.
Then we have

$$\nabla_{Y_w}X(c(t))=\sum_l\left(\sum_{i,j}X_i(c(t))Y_j(c(t))\Gamma_{ij}^l(c(t))+dX_l(c(t))Y_w(c(t))\right)\frac{\partial}{\partial\varphi_l}\Big|_{c(t)},$$

where $\Gamma_{ij}^l$ denote the Christoffel symbols of $\nabla$ on $\mathcal{B}_R(p)$ for $i,j,l\in\{1,...,k\}$, cf. (2.2).
Analogously, for a vector field $Z(p'):=\sum_{i=1}^kZ_i(p')\frac{\partial}{\partial\varphi_i}\Big|_{p'}$,

$$\nabla_{\dot{c}} Z(c(t)) = \sum_l \left( \sum_{i,j} C_i(c(t)) Z_j(c(t)) \Gamma^l_{ij}(c(t)) + dZ_l(c(t))\dot{c} \right) \frac{\partial}{\partial \varphi_l}\Big|_{c(t)}.$$

Combining these formulas shows that

$$\nabla_{\dot{c}} \nabla_{Y_w} X(c(t)) = \sum_l d_l(c(t)) \frac{\partial}{\partial \varphi_l}\Big|_{c(t)},$$

where

$$\begin{aligned} d_l(c(t)) = &\sum_{i,j,r,s} C_i(c(t)) X_r(c(t)) Y_s(c(t)) \Gamma^l_{ij}(c(t)) \Gamma^j_{rs}(c(t)) + \\ &\sum_{i,j} C_i(c(t)) dX_j(c(t)) Y_w(c(t)) \Gamma^l_{ij}(c(t)) + \\ \sum_{r,s} &\left( \frac{d}{dt}(X_r(c(t))) Y_s(c(t)) \Gamma^l_{rs}(c(t)) + X_r(c(t)) \frac{d}{dt}(Y_s(c(t))) \Gamma^l_{rs}(c(t)) + \right. \\ &\left. X_r(c(t)) Y_s(c(t)) \frac{d}{dt}(\Gamma^l_{rs}(c(t))) \right) + \frac{d}{dt}(dX_l(c(t)) Y_w(c(t))). \end{aligned}$$

We use the following abbreviations

1. $|C| := \max\{|C_i(c(t))| : t \in [0,1], i = 1, ..., k\}$,
2. $|X| := \max\{|X_i(c(t))| : t \in [0,1], i = 1, ..., k\}$,
3. $|dX| := \max\{|dX_i(c(t))v| : t \in [0,1], v \in T_{c(t)}M, \|v\|_{c(t)} = 1, i = 1, ..., k\}$,
4. $|Y| := \max\{|Y_i(c(t))| : t \in [0,1], i = 1, ..., k\}$,
5. $|\Gamma| := \max\{|\Gamma^l_{ij}(c(t))| : t \in [0,1], i, j, l = 1, ..., k\}$,
6. $|d\Gamma| := \max\{|d\Gamma^l_{ij}(c(t)) \frac{\dot{c}(t)}{\|\dot{c}(t)\|_{c(t)}}| : t \in [0,1], i, j, l = 1, ..., k\}$,
7. $|d^2X| := \max\{|H_{X_i}(c(t))(v, w)| : t \in [0,1], v, w \in T_{c(t)}M, \|v\|_{c(t)} = \|w\|_{c(t)} = 1, i = 1, ..., k\}$,
8. $\left|\frac{\partial}{\partial \varphi}\right| := \max\left\{ \left\| \frac{\partial}{\partial \varphi_i}\Big|_{c(t)} \right\|_{c(t)} : \ t \in [0,1], \ i = 1, ..., k \right\}$.

We use the following estimates

i) $dX_j(c(t))Y_w(c(t)) = \langle \operatorname{grad} X_j(c(t)), Y_w(c(t)) \rangle_{c(t)} \leq (|d^2X| \|\dot{c}(0)\|_p + |dX|) \|w\|_q$

   This can be seen by noting that

$$dX_j(c(t))Y_w(c(t)) = dX_j(p)Y_w(p) + \int_0^1 \frac{d}{dt}(dX_j(c(rt))Y_w(c(rt)))dr.$$

ii) $\frac{d}{dt}(dX_l(c(t))Y_w(c(t))) = \langle H_{X_l}(c(t))\dot{c}, Y_w(c(t))\rangle_{c(t)} + \langle \text{grad}\, X_l(c(t)), \underbrace{\frac{D}{dt}Y_w(c(t))}_{=0}\rangle_{c(t)}$

$\leq |d^2X|\|\dot{c}(0)\|_p\|w\|_q$

iii) $\frac{d}{dt}(Y_s(c(t))) = -\sum_{rl} C_r Y_l \Gamma^s_{rl}$, since

$$0 = \nabla_{\dot{c}}Y_w = \sum_s \frac{d}{dt}(Y_s(c(t)))\frac{\partial}{\partial\varphi_s}\Big|_{c(t)} + \sum_{rl} Y_l C_r \Gamma^s_{rl}\frac{\partial}{\partial\varphi_s}\Big|_{c(t)}$$

Then we get

$$|d_l| \leq k^4|C||X||Y||\Gamma|^2 + k^2|C||\Gamma|(|dX| + |d^2X|\|\dot{c}(0)\|_p)\|w\|_q + k^2|Y||\Gamma||dX|\|\dot{c}(0)\|_p +$$
$$k^4|X||Y||C||\Gamma|^2 + k^2|X||Y||d\Gamma|\|\dot{c}(0)\|_p + |d^2X|\|\dot{c}(0)\|_p\|w\|_q,$$

for $l = 1, ..., k$.
Thus we can estimate the covariant derivative of the vector field $\nabla_{Y_w}X(c(t))$ w.r.t. $\dot{c}$ as follows

$$\|\nabla_{\dot{c}}\nabla_{Y_w}X(c(t))\|_{c(t)} \leq k\left|\frac{\partial}{\partial\varphi}\right|\big(k^4|C||X||Y||\Gamma|^2 + k^2|C||\Gamma|(|dX| + |d^2X|\|\dot{c}(0)\|_p)\|w\|_q +$$
$$k^2|Y||\Gamma||dX|\|\dot{c}(0)\|_p + k^4|X||Y||C||\Gamma|^2 + k^2|X||Y||d\Gamma|\|\dot{c}(0)\|_p + |d^2X|\|\dot{c}(0)\|_p\|w\|_q\big).$$

Let $\lambda_1(t), ..., \lambda_k(t)$ denote the eigenvalues of the gramian

$$G(t) = \left(\left\langle \frac{\partial}{\partial\varphi_i}\Big|_{c(t)}, \frac{\partial}{\partial\varphi_j}\Big|_{c(t)}\right\rangle_{i,j=1}^k\right)$$

and define

$$m := \min\{\lambda_i(t) :\ t \in [0,1], i = 1, ..., k\}.$$

Then we get

$$|C| \leq \frac{\|\dot{c}(t)\|_{c(t)}}{m} = \frac{\|\dot{c}(0)\|_p}{m}$$

and

$$|Y| \leq \frac{\|w\|_q}{m},$$

which shows that there exists a constant $\kappa > 0$, depending on $m$, $|\Gamma|$, $|\Gamma|^2$, $|d\Gamma|$, $|X|$, $|dX|$ and $|d^2X|$, such that

$$\|\nabla_{\dot{c}}\nabla_{Y_w}X(c(t))\|_{c(t)} \leq \kappa\|\dot{c}(0)\|_p\|w\|_q.$$

Here,

$$\kappa \leq k\left|\frac{\partial}{\partial\varphi}\right| \cdot \Bigg(\frac{k^4}{m^2}|X||\Gamma|^2 +$$
$$\frac{k^2}{m}|\Gamma|(|dX| + |d^2X|\|\dot{c}(0)\|_p) + \frac{k^2}{m}|\Gamma||dX| + \frac{k^4}{m^2}|X||\Gamma|^2 + \frac{k^2}{m}|X||d\Gamma| + |d^2X|\Bigg)$$

**Corollary 2.1.** *Let $M$ be a connected and complete Riemannian manifold and let $X : M \to TM$ a smooth vector field. Then for $0 < R \leq i_p^*(M)$, there exists a chart $\varphi$ such that the vector field $X\big|_{\mathcal{B}_R(p)}$ can be expressed by*

$$X(p') := \sum_{i=1}^{k} X_i(p') \frac{\partial}{\partial \varphi_i}\Big|_{p'},$$

*where $\frac{\partial}{\partial \varphi_1}\Big|_{p'}, ..., \frac{\partial}{\partial \varphi_k}\Big|_{p'}$ is the coordinate frame associated with $\varphi$ and $X_i : \mathcal{B}_R(p) \to \mathbb{R}$ is a smooth function for $i = 1, ..., k$.*
*Assume that the following suprema are finite*

*(A1)* $|X| := \sup\{|X_i(p')| : p' \in \mathcal{B}_R(p), i = 1, ..., k\}$,

*(A2)* $|dX| := \sup\{|dX_i(p')v| : p' \in \mathcal{B}_R(p), v \in T_{p'}M, \|v\|_{p'} = 1, i = 1, ..., k\}$,

*(A3)* $|d^2X| := \sup\{|H_{X_i}(p')(v, w)| : p' \in \mathcal{B}_R(p), v, w \in T_{p'}M, \|v\|_{p'} = \|w\|_{p'} = 1, i = 1, ..., k\}$,

*(A4)* $|\Gamma| := \sup\{|\Gamma_{ij}^l(p')| : p' \in \mathcal{B}_R(p), i, j, l = 1, ..., k\}$,

*(A5)* $|d\Gamma| := \sup\{|d\Gamma_{ij}^l(p')v|\}| : p' \in \mathcal{B}_R(p), v \in T_{p'}M, \|v\|_{p'} = 1,\ i, j, l = 1, ..., k\}$,

*(A6)* $\left|\frac{\partial}{\partial \varphi}\right| := \sup\left\{\left\|\frac{\partial}{\partial \varphi_i}\Big|_{p'}\right\|_{p'} :\ p' \in \mathcal{B}_R(p), i = 1, ..., k\right\}$,

*(A7)* $m := \min\{\lambda_i(t) :\ t \in [0, 1], i = 1, ..., k\}$, *where $\lambda_1(t), ..., \lambda_k(t)$ denote the eigenvalues of the gramian* $G(t) = \left(\left\langle \frac{\partial}{\partial \varphi_i}\Big|_{c(t)}, \frac{\partial}{\partial \varphi_j}\Big|_{c(t)} \right\rangle_{i,j=1}^{k}\right)$

*Then the following statements hold:*

1. *There exists a constant $\kappa > 0$ such that for $q \in \mathcal{B}_R(p)$ and $w \in T_qM$*

$$\|\nabla_{\dot{c}} \nabla_{Y_w} X(c(t))\|_{c(t)} \leq \kappa\ \|\dot{c}(0)\|_p \|w\|_q, \tag{2.8}$$

   *where $Y_w : \mathcal{B}_R(p) \to TM$ is defined by $Y_w(p') = \tau_{pp'}(\tau_{pq}^{-1} w)$.*

2. *If $M$ is compact and connected, then there exists $r > 0$ and $\kappa > 0$ such that (2.8) holds uniformly for all $p \in M$, $q \in \mathcal{B}_r(p)$ and $w \in T_qM$.*

This shows that one is able to bound higher order covariant derivatives by local expressions of the vector field $X$ via coordinate frames and the Christoffel symbols. From now on, we assume the existence of such a bound on the covariant derivatives in order to estimate the norm of the vector field.

**Lemma 2.3.** *Let $M$ be a complete Riemannian manifold, $X \in \mathcal{V}^\infty(M)$, $p \in M$ and $R = i_p^*(M)$ and let $(c_q)_{q \in \mathcal{B}_R(p)}$ denote the family of geodesic curves from $p$ to $q$ with $c_q : [0, R_q] \to M$ and $R_q = \operatorname{dist}(p, q) \leq R$ for all $q \in \mathcal{B}_R(p)$.*
*Assume the existence of constants $c_1, c_2, c_3 > 0$ such that*

1. *$c_1 \leq \|\nabla_v X(p)\|_p \leq c_2$, for all $v \in T_pM$ with $\|v\|_p = 1$.*
2. *$\|\nabla_{\dot{c}_q} \nabla_{Y_w} X(c_q(t))\|_{c_q(t)} \leq c_3$ for all $q \in \mathcal{B}_R(p)$, $t \in [0, R_q]$ and $w \in T_qM$ with $\|w\|_q = 1$. Here $Y_w$ is a smooth vector field on $\mathcal{B}_R(p)$ defined by $Y_w(p') := \tau_{pp'}\tau_{pq}^{-1} w$.*

*Then we get for $r \leq \min\{R, \frac{c_1}{2c_3}\}$ that*

$$\frac{c_1}{2}\operatorname{dist}(p, q) \leq \|\tau_{pq}^{-1} X(q) - X(p)\|_p \leq (c_1/2 + c_2)\operatorname{dist}(p, q) \tag{2.9}$$

*for all $q \in \mathcal{B}_r(p)$. Moreover, the covariant derivative of $X$ at $q \in \mathcal{B}_r(p)$ with respect to $w \in T_qM$ with $\|w\|_q = 1$ satisfies*

$$\frac{c_1}{2} \leq \|\nabla_w X(q)\|_q \tag{2.10}$$

*Proof.* Let $d(p, q) \leq r$ and let $\hat{v} \in T_pM$ such that $q = \exp_p \hat{v}$. Then the geodesic curve $c_q : [0, R_q] \to M$ from $p$ to $q$ with $R_q = \operatorname{dist}(p, q)$ and $\|\dot{c}_q\|_{c_q(t)} = 1$ is given by $c_q(t) := \exp_p(t\hat{v}/\|\hat{v}\|_p)$. According to (2.6), we have for $n = 2$

$$\tau_{pq}^{-1} X(q) = X(p) + R_q \nabla_{\dot{c}_q(0)} X(p) + R_q^2 \int_0^1 (1 - s)\tau_{pc_q(R_q s)}^{-1} \nabla^2_{\dot{c}_q(R_q s)} X(c_q(R_q s)) ds. \tag{2.11}$$

Note that for $w := \dot{c}_q(R_q)$ and $0 \leq s \leq 1$ we have $Y_w(c_q(R_q s)) = \tau_{pc_q(R_q s)}\tau_{pq}^{-1}\dot{c}_q(R_q) = \tau_{pc_q(R_q s)}\dot{c}_q(0) = \dot{c}_q(R_q s)$. Thus, the right side of (2.9) gets obvious by noting that

$$\left\| R_q \nabla_{\dot{c}_q(0)} X(p) + R_q^2 \int_0^1 (1 - s)\tau_{pc_q(R_q s)}^{-1} \nabla^2_{\dot{c}_q(R_q s)} X(c_q(R_q s)) ds \right\|_p \leq$$

$$\operatorname{dist}(p, q)(c_2 + \operatorname{dist}(p, q)c_3) \leq \operatorname{dist}(p, q)(c_2 + c_1/2),$$

where we used $\|\nabla^2_{\dot{c}_q(R_q s)} X(c_q(R_q s))\|_{c_q(R_q s)} = \|\nabla_{\dot{c}_q(R_q s)} \nabla_{Y_w} X(c_q(R_q s))\|_{c_q(R_q s)} \leq c_3$ for $w = \dot{c}_q(R_q)$, due to assumption (2).
On the other hand, (2.11) implies that

$$\|\tau_{pq}^{-1} X(q) - X(p)\|_p \geq \|R_q \nabla_{\dot{c}_q(0)} X(p)\|_p - \left\| R_q^2 \int_0^1 (1 - s)\tau_{pc_q(R_q s)}^{-1} \nabla^2_{\dot{c}_q(R_q s)} X(c_q(R_q s)) ds \right\|_p$$

Since $\operatorname{dist}(p, q) = R_q \leq r \leq \frac{c_1}{2c_3}$, we get

$$\|\tau_{pq}^{-1} X(q) - X(p)\|_p \geq \operatorname{dist}(p, q)(c_1 - rc_3).$$

And therefore

$$\|\tau_{pq}^{-1}X(q) - X(p)\|_p \geq \text{dist}(p,q)c_1/2.$$

Equation 2.10 can be seen by applying Taylor's formula to the vector field $\nabla_{Y_w}X$, where $Y_w(p') := \tau_{pp'}\tau_{pq}^{-1}w$, $p' \in \mathcal{B}_r(p)$. Thus,

$$\tau_{pq}^{-1}\nabla_w X(q) = \nabla_{Y_w}X(p) + R_q \int_0^1 \tau_{pc_q(R_q s)}^{-1}\nabla_{\dot{c}(R_q s)}\nabla_{Y_w}X(c_q(R_q s))ds,$$

which shows that

$$\|\nabla_w X(q)\|_q \geq \|\nabla_{Y_w}X(p)\|_p - R_q c_3 \geq c_1 - c_1/2.$$

□

### 2.1.2 The tracking algorithms

We now have the necessary tools to analyze the time-varying Newton flow on Riemannian manifolds. This general approach is the basis of the subsequent chapters.
Let $M$ be a connected and complete Riemannian manifold. We consider a smooth time-varying vector field

$$F : M \times \mathbb{R} \to TM,$$

defined by

$$(x,t) \mapsto F(x,t) \in T_x M.$$

The object of interest is a smooth zero of $F$, which is a smooth curve $x_* : \mathbb{R} \to M$, satisfying $F(x_*(t),t) = 0$ for all $t$. This smooth zero is called isolated, if there exists a $r > 0$ such that for all $t \in \mathbb{R}$ and $x \in \mathcal{B}_r(x_*(t))$ holds: $F(x,t) = 0$ if and only if $x = x_*(t)$.
In order to determine $x_*(t)$, we consider derivatives of the vector field $F$. As usual, we use the notation $\nabla_{\dot{x}}F(x,t)$ representing the covariant derivative of $F$ on $M$ with respect to $\dot{x} \in T_x M$ and the affine connection $\nabla$ for $(x,t) \in M \times \mathbb{R}$.
To derive a differential equation on $M$, whose solution asymptotically tracks the smooth zero of $F$, we further need the following concept.

**Definition 2.3.** *Let $\mathcal{M} : TM \to TM$ a smooth bundle map, defined by linear maps $\mathcal{M}(x) : T_x M \to T_x M$ for $x \in M$. $\mathcal{M}$ is called stable, if there exists $b > 0$ such that*

$$\langle \mathcal{M}(x) \cdot v, v \rangle_x \leq -b,$$

*for all $x \in M$ and $v \in T_x M$ with $\|v\|_x = 1$.*

Examples for stable bundle maps are given by $\mathcal{M}(x) = -\sigma I$ for $\sigma > 0$.

Now consider the ODE

$$\boxed{\nabla_{\dot{x}}F(x,t) + \frac{\partial}{\partial t}F(x,t) = \mathcal{M}(x)F(x,t),} \tag{2.12}$$

and note that the left hand side of (2.12) equals $\frac{D}{dt}F(x(t),t)$, cf. Lemma 2.1. We call the differential equation (2.12) *time-varying Newton flow*, as the Newton flow is a special case of this equation in the time-invariant case, where $F$ does not depend explicitly on $t$.
It will be shown in the proof of the following theorem, that solutions $x(t)$ of (2.12) satisfy

$$\|F(x(t),t)\|_{x(t)} \leq ae^{-bt}$$

for some $a, b > 0$ and all $t > 0$. However, this alone does not imply that $x(t)$ converges to the zero $x_*(t)$ of $F$, which can be easily seen by considering the example $F(x,t) := xe^{-t}$ for $M = \mathbb{R}$.
A further problem is, that the differential equation (2.12) is in an implicit form. To overcome this, we need conditions which guarantee, that the covariant derivative $\nabla F(x,t) : T_xM \to T_xM$ is invertible in a neighborhood of $x_*(t)$, i.e. there exists a $r > 0$ such that for $t \in \mathbb{R}$

$$\mathrm{rk}(\nabla F(x,t)) = \dim M, \quad \forall x \in \mathcal{B}_r(x_*(t)).$$

Then (2.12) can be rewritten in explicit form

$$\boxed{\dot{x} = (\nabla F(x,t))^{-1}\left(\mathcal{M}(x)F(x,t) - \frac{\partial}{\partial t}F(x,t)\right).} \tag{2.13}$$

These examinations motivate the use of additional assumptions on $F$ in our first main result.

**Main Theorem 2.1.** *Let $M$ be a complete Riemannian manifold and $R > 0$ any real number with $i^*(M) \geq R$. Let $F : M \times \mathbb{R} \to TM$ be a smooth time-varying vector field on $M$ and let $t \mapsto x_*(t)$ be a smooth isolated zero of $F$, i.e. $F(x_*(t),t) = 0$ for all $t \in \mathbb{R}$. Assume further the existence of constants $c_1, c_2, c_3 > 0$ such that the following conditions are satisfied for all $t \in \mathbb{R}$*

*(i)* $c_1 \leq \|\nabla_v F(x_*(t),t)\|_{x_*(t)} \leq c_2$ *for* $v \in T_{x_*(t)}M$ *with* $\|v\|_{x_*(t)} = 1$.

*(ii)* $\|\nabla_{\dot{c}_x(s)}\nabla_{Y_w}F(c_x(s),t)\|_{c_x(s)} \leq c_3$ *for all* $x \in \mathcal{B}_R(x_*(t))$, $s \in [0,R_x]$ *and* $w \in T_xM$ *with* $\|w\|_x = 1$.

*Here* $R_x := \mathrm{dist}(x_*(t),x)$ *and* $c_x : [0,R_x] \to M$ *is a geodesic from* $x_*(t)$ *to* $x$ *with* $\|\dot{c}_x(s)\|_{c_x(s)} = 1$ *for* $s \in [0,R_x]$, *defined by* $c_x(s) = \exp_{x_*(t)}\left(s\frac{\exp^{-1}_{x_*(t)}(x)}{\|\exp^{-1}_{x_*(t)}(x)\|_{x_*(t)}}\right)$ *and* $Y_w$ *is a smooth vector field on* $\mathcal{B}_R(x_*(t))$ *defined by* $Y_w(x') := \tau_{x_*(t)x'}\tau^{-1}_{x_*(t)x}w$.

*Then for* $r \leq \min\{R, \frac{c_1}{2c_3}\}$, *the linear map* $\nabla F(x,t) : T_xM \to T_xM$ *is an isomorphism for all* $x \in \mathcal{B}_r(x_*(t))$, $t \in \mathbb{R}$. *Moreover, for all* $x \in \mathcal{B}_r(x_*(t))$ *and* $t \in \mathbb{R}$

$$\frac{c_1}{2}\mathrm{dist}(x,x_*(t)) \leq \|F(x,t)\|_x \leq (c_1/2 + c_2)\mathrm{dist}(x,x_*(t)). \tag{2.14}$$

*Thus, if $x(0) \in M$ is sufficiently close to $x_*(0)$ and $\mathcal{M}$ is a stable bundle map, the solution $x(t) \in M$ of (2.13) converges exponentially to $x_*(t)$, i.e. for some $a, b > 0$ holds for all $t \in \mathbb{R}$*

$$\operatorname{dist}(x(t), x_*(t)) \leq ae^{-bt}.$$

*Proof.* Note that for $t \in \mathbb{R}$ equation (2.14) already has been shown in Lemma 2.3. We moreover get, that the smallest singular value of the covariant derivative is lower bounded by $c_1/2$ on $\mathcal{B}_r(x_*(t))$.
Thus, solutions $x(t)$ of (2.13) exist for $x(0) \in \mathcal{B}_r(x_*(0))$ at least locally for $t \in [0, t_m]$ and the norm estimate of $F(x(t), t)$ holds:

$$\frac{d}{dt}\|F(x(t),t))\|^2_{x(t)} =$$

$$\frac{d}{dt}\left\langle F(x(t),t), F(x(t),t)\right\rangle_{x(t)} =$$

$$2\left\langle \nabla_{\dot{x}} F(x(t),t) + \frac{\partial}{\partial t} F(x(t),t), F(x(t),t)\right\rangle_{x(t)} =$$

$$2\left\langle \mathcal{M}(x)F(x(t),t), F(x(t),t)\right\rangle_{x(t)} \leq -2b\|F(x(t),t)\|^2_{x(t)},$$

for some $b > 0$, cf. Definition 2.3. This implies that

$$\|F(x(t),t)\|_x \leq e^{-bt}\|F(x(0),0)\|_{x(0)}.$$

To guarantee the global existence of such a solution $x(t)$, we have to show that $x(t)$ remains close to $x_*(t)$: Assume, that there exists a $t_m > 0$ such that $\operatorname{dist}(x(t_m), x_*(t_m)) > r$.
If we choose $x_0$ such that $\operatorname{dist}(x(0), x_*(0)) \leq r\frac{c_1}{2(c_1/2+c_2)}$, then

$$\|F(x(t_m),t_m)\|_{x(t_m)} \leq \|F(x(0),0)\|_{x(t_m)} \leq (c_1/2 + c_2)\operatorname{dist}(x(0), x_*(0)) \leq \frac{rc_1}{2},$$

cf. (2.14). Thus, again using (2.14) shows that $\operatorname{dist}(x(t_m), x_*(t_m)) \leq r$, contradicting the assumption. □

#### 2.1.2.1 Time-varying Riemannian Newton algorithm

In order to deduce an update scheme from the explicit time-varying Newton flow (2.13), we employ a standard numerical discretization; see e.g. Stoer and Bulirsch [64] for details.
Consider the ODE

$$\dot{x} = g(x,t), \tag{2.15}$$

where $g : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ and let $t_k = kh$ for step size $h > 0$ and $k \in \mathbb{N}$. A single-step discretization of (2.15) is given by the rule:

$$x_{k+1} = x_k + h\Phi(x_k, t_k, h),$$

where $\Phi$ is any map, chosen such that $(x_k)$ is an approximation of $(x(t_k))$. A possible choice for $\Phi$ is the familiar first order Euler method, defined in Euclidean space by

$$x_{k+1} = x_k + hg(x_k, t_k). \tag{2.16}$$

In our situation, we replace the Euler discretization (2.16) by its intrinsic variant

$$x_{k+1} = \exp_{x_k}(hg(x_k, t_k)), \tag{2.17}$$

where $\exp_{x_k}$ denotes the exponential map of $M$ at $x_k \in M$.
As we want to establish a realistic update scheme, we have to replace $\frac{\partial}{\partial t}F(x,t)$ by a step size-dependent approximation $F_\tau^h(x,t)$ at discrete times $t_k = kh$, $k \in \mathbb{N}$ for some $h > 0$. How such approximations can be found is shown in Section 2.1.2.2.
Then the Riemannian update scheme (2.17) corresponding to the differential equation (2.13), is given for $\mathcal{M}(x) = -\frac{1}{h}I$ by

$$\boxed{x_{k+1} = \exp_{x_k}\left((\nabla F(x_k, t_k))^{-1}\left(-F(x_k, t_k) - hF_\tau^h(x_k, t_k)\right)\right),} \tag{2.18}$$

where $\exp_{x_k}$ denotes the exponential map of $M$ at $x_k \in M$. We call this formula the *time-varying Newton algorithm.*
The next theorem gives conditions, which guarantee that the resulting sequence $(x_k)$ is a good approximation of the smooth isolated zero $x_*(t)$ of $F(x,t)$ on $M$ at discrete times $t = t_k$.

**Main Theorem 2.2.** *Let $M$ be a complete Riemannian manifold and $R > 0$ any real number with $i^*(M) \geq R$. Let $F : M \times \mathbb{R} \to TM$, $(x,t) \mapsto F(x,t) \in T_xM$ a smooth time-varying vector field and let $t \mapsto x_*(t)$ be a smooth isolated zero of $F$, i.e. $F(x_*(t),t) = 0$ for all $t \in \mathbb{R}$. Assume further the existence of constants $c_1, c_2, c_3, c_4, c_5, c_6, c_7 > 0$ such that the following conditions are satisfied for all $t \in \mathbb{R}$*

*(i)* $\|\nabla F(x_*(t),t)\|_{x_*(t)} \leq c_1$, $\|\frac{\partial}{\partial t}F(x_*(t),t)\|_{x_*(t)} \leq c_2$, $\|\nabla F(x_*(t),t)^{-1}\|_{x_*(t)} \leq c_3$,

*(ii)* $\|\nabla_{\dot{c}_{x_0}^x(s)} \nabla_{Y_w} F(c_{x_0}^x(s),t)\|_{c_{x_0}^x(s)} \leq c_4$ *for all* $x_0, x \in \mathcal{B}_R(x_*(t))$, $s \in [0, R_{x_0}^x]$ *and* $w \in T_xM$ *with* $\|w\|_x = 1$.
*Here* $R_{x_0}^x := \operatorname{dist}(x_0, x)$ *and* $c_{x_0}^x : [0, R_{x_0}^x] \to M$ *is a geodesic from* $x_0$ *to* $x$, *defined by* $c_{x_0}^x(s) = \exp_{x_0}\left(s\frac{\exp_{x_0}^{-1}(x)}{\|\exp_{x_0}^{-1}(x)\|_{x_0}}\right)$ *and* $Y_w$ *is a smooth vector field on* $\mathcal{B}_R(x_*(t))$ *defined by* $Y_w(x') := \tau_{x_*(t)x'}\tau_{x_*(t)x}^{-1}w$.

*(iii)* $\|\frac{\partial^2}{\partial t^2}F(x,t)\|_x \leq c_5$, $\|\frac{\partial}{\partial t}\nabla F(x,t)\|_x \leq c_6$ *for all* $x \in \mathcal{B}_R(x_*(t))$.

*(iv)* $\|F_\tau^h(x,t) - \frac{\partial}{\partial t}F(x,t)\|_x \leq c_7 h$, *for all* $x \in \mathcal{B}_R(x_*(t))$, $h > 0$.

*Then the following statements hold*

1. *There exists* $0 < r \leq R$ *and* $c_8, c_9, c_{10} > 0$ *such that for* $t \in \mathbb{R}$

$$\operatorname{dist}(x, x_*(t)) \leq c_8 \|F(x,t)\|_x, \tag{2.19}$$

$$\|F(x,t)\|_x \leq c_9 \operatorname{dist}(x, x_*(t)), \tag{2.20}$$

   *and*

$$\|\nabla F(x,t)^{-1}\|_x \leq c_{10} \tag{2.21}$$

   *for* $x \in \mathcal{B}_r(x_*(t))$.

2. *The discretization sequence* $(x_k)$ *as defined in* (2.18) *with* $t_k = kh$, $h > 0$ *satisfies for some* $c_{11}, c_{12} > 0$

$$\operatorname{dist}(x_{k+1}, x_*(t_{k+1})) \leq c_{11} \operatorname{dist}(x_k, x_*(t_k))^2 + c_{12} h^2 \tag{2.22}$$

   *for* $x_k \in \mathcal{B}_r(x_*(t_k))$, $k \in \mathbb{N}_0$.

3. *Let* $c > 0$ *be constant and* $h > 0$ *sufficiently small. For any initial condition* $x_0$ *with* $\operatorname{dist}(x_0, x_*(0)) < ch$, *the sequence* (2.18) *satisfies*

$$\operatorname{dist}(x_k, x_*(t_k)) \leq ch$$

   *for all* $k \in \mathbb{N}_0$. *Thus, the update scheme* (2.18) *is well defined and produces estimates for* $x_*(t_k)$, *whose accuracy can be controlled by the step size.*

*Proof.* 1) These claims are direct consequences from Lemma 2.3.

2) Consider first the following formula, cf. equation (2.7),

$$\tau_{x_k x_{k+1}}^{-1} F(x_{k+1}, t_{k+1}) = F(x_k, t_k) + \frac{\partial F}{\partial t}(x_k, t_k)h + \nabla F(x_k, t_k)w + \mathcal{R}, \tag{2.23}$$

where $\tau_{x_k x_{k+1}}$ denotes the parallel transport along the geodesic from $x_k$ to $x_{k+1}$ and $\mathcal{R}$ satisfies

$$\|\mathcal{R}\|_{x_k} \leq c_5 h^2 + c_4 \|w\|_{x_k}^2 + c_6 h \|w\|_{x_k} \leq \left(c_4 + \frac{c_6}{2}\right) \|w\|_{x_k}^2 + \left(c_5 + \frac{c_6}{2}\right) h^2.$$

The update scheme (2.18) shows that $w = \exp_{x_k}^{-1}(x_{k+1})$ is given in our situation as

$$w = h(\nabla F(x_k, t_k))^{-1} \left(-\frac{1}{h} F(x_k, t_k) - F_\tau^h(x_k, t_k)\right).$$

Hence, equation (2.23) turns into

$$\tau_{x_k x_{k+1}}^{-1} F(x_{k+1}, t_{k+1}) = \frac{\partial F}{\partial t}(x_k, t_k)h - hF_\tau^h(x_k, t_k) + \mathcal{R}, \tag{2.24}$$

which implies

$$\|\tau_{x_k x_{k+1}}^{-1} F(x_{k+1}, t_{k+1})\|_{x_k} = \left\| \frac{\partial F}{\partial t}(x_k, t_k)h - hF_\tau^h(x_k, t_k) + \mathcal{R} \right\|_{x_k},$$

and therefore

$$\|F(x_{k+1}, t_{k+1})\|_{x_k} \leq c_7 h^2 + \left(c_4 + \frac{c_6}{2}\right) \|w\|_{x_k}^2 + \left(c_5 + \frac{c_6}{2}\right) h^2. \qquad (2.25)$$

Now consider

$$\|w\|_{x_k} = \left\|(\nabla F(x_k, t_k))^{-1} \left(F(x_k, t_k) + hF_\tau^h(x_k, t_k)\right)\right\|_{x_k},$$

showing that

$$\|w\|_{x_k} \leq c_{10}(\|F(x_k, t_k)\|_{x_k} + h\|F_\tau^h(x_k, t_k)\|_{x_k}). \qquad (2.26)$$

Note that $\|F_\tau^h(x_k, t_k)\|_{x_k} \leq \|\frac{\partial}{\partial t} F(x_k, t_k)\|_{x_k} + c_7 h \leq c_2 + c_6 \text{dist}(x_k, x_*(t_k)) + c_7 h$. Thus (2.26) turns into

$$\|w\|_{x_k} \leq c_{10}\big(\|F(x_k, t_k)\|_{x_k} + h(c_2 + c_6 \text{dist}(x_k, x_*(t_k)) + c_7 h)\big),$$

implying that

$$\|w\|_{x_k} \leq c_{10}\big(c_9 \text{dist}(x_k, x_*(t_k)) + h(c_2 + c_6 r + c_7 h)\big).$$

By using the abbreviations $k_1 = c_{10} c_9$ and $k_2 = c_{10}(c_2 + c_6 r + c_7 h)$, this equation shows that

$$\|w\|_{x_k}^2 \leq k_1^2 \text{dist}(x_k, x_*(t_k))^2 + k_2^2 h^2 + 2k_1 k_2 \text{dist}(x_k, x_*(t_k)) h,$$

and hence

$$\|w\|_{x_k}^2 \leq (k_1^2 + k_1 k_2) \text{dist}(x_k, x_*(t_k))^2 + (k_2^2 + k_1 k_2) h^2.$$

Plug this into (2.25) and obtain

$$\|F(x_{k+1}, t_{k+1})\|_{x_k} \leq k_3(k_1^2 + k_1 k_2) \text{dist}(x_k, x_*(t_k))^2 + \left(k_3 k_2^2 + k_1 k_2 k_3 + c_7 + c_5 + \frac{c_6}{2}\right) h^2,$$

where $k_3 = (c_4 + c_6/2)$. Using (2.19) shows the claim for $c_{11} = c_8 k_3 (k_1^2 + k_1 k_2)$ and $c_{12} = c_8(k_3 k_2^2 + k_1 k_2 k_3 + c_7 + c_5 + \frac{c_6}{2})$.

3) Suppose

$$\text{dist}(x_k, x_*(t_k)) \leq ch$$

holds for some $k$. By the estimate (2.22) then

$$\text{dist}(x_{k+1}, x_*(t_{k+1})) \leq$$

$$c_{11} \text{dist}(x_k, x_*(t_k))^2 + c_{12} h^2 \leq c_{11} c^2 h^2 + c_{12} h^2 \leq ch, \qquad (2.27)$$

for $h < \frac{c}{c_{11} c^2 + c_{12}}$. □

**Remark 2.3.**

1. Standard estimates for the discretization error imply the above result 3) only for a finite number of iterates $x_k$. The new interesting feature therefore is that the error estimate can be guaranteed to hold for all $k \in \mathbb{N}_0$. This is due to the fact, that we have chosen $\mathcal{M}(x) = -\frac{1}{h}I$. Without that choice we would not be able to prove a similar estimate.

2. In the special case that $F$ is the gradient of a function $f : M \times \mathbb{R} \to \mathbb{R}$, i.e.

$$F(x,t) = \operatorname{grad} f(x,t),$$

the update scheme (2.18) turns into

$$x_{k+1} = \exp_{x_k}\left(-H_f(x_k,t_k)^{-1}\left(\operatorname{grad} f(x_k,t_k) + hG_h(x_k,t_k)\right)\right), \tag{2.28}$$

where $H_f$ denotes the Riemannian Hesse operator and $G_h(x,t)$ denotes an approximation of $\frac{\partial}{\partial t}\operatorname{grad} f(x,t)$.

Note that under the conditions of the previous theorem, the tracking property of the update scheme defined in (2.18) even remains, if one uses the 0th order approximation $F_\tau^h(x,t) = 0$ for $\frac{\partial}{\partial t}F(x,t)$. This shows, that the conventional Riemannian Newton algorithm can be used to track the zero of a time-varying vector field. We formulate this in the following corollary.

**Corollary 2.2.** *Assume (i) – (iii) of Theorem 2.2 and*

*(iv') $\left\|\frac{\partial}{\partial t}F(x,t)\right\|_x \leq c_7$ for some $c_7 > 0$ and all $x \in \mathcal{B}_R(x_*(t))$, $t \in \mathbb{R}$.*

*Then the Newton update scheme*

$$x_{k+1} = \exp_{x_k}\left((\nabla F(x_k,t_k))^{-1}(-F(x_k,t_k))\right)$$

*satisfies the weak tracking property, i.e. for any $0 < \hat{r} \leq r$ there exists $h > 0$ and $0 < r_0 \leq \hat{r}$ such that $\operatorname{dist}(x_k, x_*(t_k)) \leq \hat{r}$ for all $k \in \mathbb{N}$, provided $\operatorname{dist}(x_0, x_*(0)) \leq r_0$.*

*Proof.* Note first, that with the notation of Theorem 2.2, the following equation holds instead of (2.22)

$$\operatorname{dist}(x_{k+1}, x_*(t_{k+1})) \leq c_{11}\operatorname{dist}(x_k, x_*(t_k))^2 + c_{12}h,$$

which can be immediately seen by considering the proof of the above theorem.
Thus it suffices to show, that for $a, b, R > 0$ and a real valued sequence $(y_k)_{k\in\mathbb{N}_0}$ satisfying for $k \in \mathbb{N}$

$$0 \leq y_{k+1} \leq a y_k^2 + bh, \tag{2.29}$$

there exists $0 < r < R$ such that $y_k \leq r$ for all $k \in \mathbb{N}$, provided that $0 \leq y_0 \leq r$ and $h > 0$ is sufficiently small.
For $y_0 \geq 0$, we get that $y_k \geq 0$ for all $k \in \mathbb{N}$. Thus it remains to show, that for $h$ sufficiently small, there exists $0 < r \leq R$ and $0 < A_1 < r$ such that the following holds:

1. $y_k \in [A_1, r]$ implies that $y_{k+1} \leq y_k$,
2. If $y_k \in [0, A_1)$, then $y_{k+1} \leq r$.

Then the sequence $(y_k)$ can not leave the interval $[0, r]$.
1) The sequence is not increasing if $y_{k+1} \leq y_k$, i.e.

$$ay_k^2 + bh \leq y_k,$$

which is equivalent to

$$ay_k^2 - y_k + bh \leq 0.$$

This equation is solvable, if $D := 1 - 4abh \geq 0$. This gives a first condition for $h$. The set of solution of this inequality is then $y_k \in [A_1, A_2] \subset \mathbb{R}_+$, where $A_1 = \frac{1-\sqrt{D}}{2a}$ and $A_2 = \frac{1+\sqrt{D}}{2a}$. Note that $A_1, A_2 > 0$ as $D \in [0, 1)$. Choose $h$ sufficiently small such that $A_1 + bh \leq R$ and $bh \leq 2\frac{\sqrt{D}}{2a}$. This is possible, as $\lim_{h\to 0} D = 1$ and $\lim_{h\to 0} A_1 = 0$. Then for $r := A_1 + bh$ holds $[A_1, r] \subset [A_1, A_2]$ and in this interval, the sequence $(y_k)$ is monotonically decreasing.
2) Compute $y_{k+1}$ for $y_k \in [0, A_1]$.

$$\begin{aligned} y_{k+1} &\leq ay_k^2 + bh \leq aA_1^2 + bh \\ &= A_1\left(\frac{1-\sqrt{D}}{2}\right) + bh \leq \frac{1}{2}A_1 + bh < A_1 + bh = r \end{aligned}$$

which shows that $y_{k+1} \in [0, r]$. □

#### 2.1.2.2 Approximations for $\frac{\partial}{\partial t} g(t)$

In order to implement the discrete algorithms one needs an exact formula for the partial derivatives $\frac{\partial F}{\partial t}(x_k, t_k)$. Often this is a restriction, as the precise values may not be available or corrupted by noise. Thus one has to replace the partial derivative by suitable higher order Taylor approximations $F_\tau^h(x, t, h)$ of $\frac{\partial F}{\partial t}(x, t)$.
In general, one finds an $m$th-order approximations for a time-varying map $g(t)$ by considering its Taylor series:

$$g(t+h) = g(t) + g'(t)h + \frac{1}{2}g''(t)h^2 + \frac{1}{6}g'''(t)h^3 + \frac{1}{24}g''''(t)h^4 + O(h^5)$$

Similarly, we develop the Taylor series at further points:

$$g(t-h) = g(t) - g'(t)h + \frac{1}{2}g''(t)h^2 - \frac{1}{6}g'''(t)h^3 + \frac{1}{24}g''''(t)h^4 + O(h^5)$$

$$g(t+2h) = g(t) + g'(t)2h + \frac{4}{2}g''(t)h^2 + \frac{8}{6}g'''(t)h^3 + \frac{16}{24}g''''(t)h^4 + O(h^5)$$

$$g(t-2h) = g(t) - g'(t)2h + \frac{4}{2}g''(t)h^2 - \frac{8}{6}g'''(t)h^3 + \frac{16}{24}g''''(t)h^4 + O(h^5)$$

| Approximation for $g'(t)$ | Order |
|---|---|
| $0$ | 0 |
| $\frac{1}{2h}\Big(g(t+h)-g(t-h)\Big)$ | 2 |
| $\frac{1}{12h}\Big(8g(t+h)-8g(t-h)-g(t+2h)+g(t-2h)\Big)$ | 4 |

Table 2.1: The order of different symmetric approximations for $g'(t)$

| Approximation for $g'(t)$ | Order |
|---|---|
| $\frac{1}{h}\Big(g(t)-g(t-h)\Big)$ | 1 |
| $\frac{1}{2h}\Big(3g(t)-4g(t-h)+g(t-2h)\Big)$ | 2 |
| $\frac{1}{6h}\Big(11g(t)-18g(t-h)+9g(t-2h)-2g(t-3h)\Big)$ | 3 |

Table 2.2: The order of different approximations for $g'(t)$

The task is now to add these maps, such that only $g'(t)h$ and some higher order terms remain. This is equivalent to solving the linear equation

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 2 & -2 \\ \frac{1}{2} & \frac{1}{2} & 2 & 2 \\ \frac{1}{6} & -\frac{1}{6} & \frac{4}{3} & -\frac{4}{3} \end{bmatrix} x = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix},$$

whose solution is $x = \frac{1}{12}(8,-8,-1,1)^\top$. Therefore, $\frac{1}{12}(8g(t+h)-8g(t-h)-g(t+2h)+g(t-2h))$ is an approximation of $g'(t)h$ of at least order 4. The resulting approximation formulas are given in Table 2.1.

We can get similar approximation formulas by only evaluating preceding points $t-kh$, $k \in \mathbb{N}_0$, in order to have a real "online algorithm", where no future data is known. Hence, we use $g(t)$, $g(t-h)$, $g(t-2h)$ and

$$g(t-3h) = g(t) - g'(t)3h + \frac{9}{2}g''(t)h^2 - \frac{27}{6}g'''(t)h^3 + \frac{81}{24}g''''(t)h^4 + O(h^5).$$

This yields the formulas of Table 2.2.

### 2.1.3 Newton flows on Riemannian submanifolds

In this section, we consider the case of $M$ being a Riemannian submanifold of $\mathbb{R}^n$. This additional assumption will lead to more explicit versions of the tracking algorithms, since one has the useful *Gauss formula for Riemannian submanifolds*, which is given in the next lemma.

**Lemma 2.4.** *Let $\nabla$ be the Levi-Civita connection for a Riemannian submanifold $(M, g_M)$ of a Riemannian manifold $(N, g)$ with induced metric $g_M = g\big|_{TM\times TM}$. Given smooth vector fields $X, Y : M \to TM$ let $\hat{X}, \hat{Y}$ be any extensions to smooth vector fields on $N$. Then the Levi-Civita connections $\hat{\nabla}$ on $N$ and $\nabla$ on $M$ are related by*

$$\nabla_X Y(x) = \pi_{T_xM} \circ \hat{\nabla}_{\hat{X}} \hat{Y}(x),$$

*for all $x \in M$. Here $\pi_{T_xM} \subset N$ denotes the projection onto the tangent space $T_xM$ of $M$ at $x$.*

We will exploit this fact to obtain simple expressions for the Levi-Civita connection on $M$. Thus, we now consider time-varying Newton flows on a smooth Riemannian submanifold $M \subset \mathbb{R}^n$, where *we always assume that $\mathbb{R}^n$ has been endowed with the Euclidean inner product* and any Riemannian submanifold $M \subset \mathbb{R}^n$ will be endowed with the induced Riemannian metric.
We consider a smooth map

$$\hat{F} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n,$$
$$(x, t) \mapsto \hat{F}(x, t)$$

and assume that its restriction

$$F := \hat{F}\big|_{M\times\mathbb{R}}$$

is a vector field, i.e. $F(x, t) \in T_xM$ for all $t \in \mathbb{R}$. We further assume, that there exists a continuous curve $x_* : \mathbb{R} \to M$ such that for all $t \in \mathbb{R}$ holds

$$F(x_*(t), t) = 0.$$

From the previous section, we get that the dynamical system to track the zero $x_*(t)$ is given by

$$\pi_{T_xM} D\hat{F}(x, t) \cdot \dot{x} + \frac{\partial}{\partial t} F(x, t) = \mathcal{M}(x) F(x, t), \tag{2.30}$$

where $\mathcal{M}$ is a stable bundle map, $\pi_{T_xM}$ denotes the projection onto the tangent space $T_xM$ and $D\hat{F}(x, t)$ is the usual derivative of $\hat{F}$ with respect to $x$ in the ambient space. Thus, $\pi_{T_xM} D\hat{F}(x, t) : T_xM \to T_xM$ is the covariant derivative of $F(x, t)$ with respect to $x \in M$, cf. Lemma 2.4. We therefore get a more concrete version of the tracking algorithm formulated in Main Theorem 2.1.

**Theorem 2.3.** *Let $M \subset \mathbb{R}^n$ be a complete Riemannian submanifold and $R > 0$ any real number with $i^*(M) \geq R$. Let $\hat{F} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$, $(x, t) \mapsto \hat{F}(x, t)$ be smooth, let $F = \hat{F}\big|_{M\times\mathbb{R}}$ with $F(x, t) \in T_xM$ and let $t \mapsto x_*(t) \in M$ be a smooth isolated zero of $F$, i.e. $F(x_*(t), t) = 0$ for all $t$. Assume further the existence of constants $c_1, c_2, c_3 > 0$ such that for all $t \in \mathbb{R}$ holds:*

1. *$c_1 \leq \|\pi_{T_{x_*(t)}M} D\hat{F}(x_*(t), t) \cdot v\| \leq c_2$ for all $v \in T_{x_*(t)}M$ with $\|v\| = 1$,*
2. *$\|\pi_{T_xM} D(\pi_{T_xM} D\hat{F}(x, t) \cdot v) \cdot w\| \leq c_3$ for all $v, w \in T_xM$ with $\|v\| = \|w\| = 1$, $x \in \mathcal{B}_R(x_*(t))$.*

*Then the solution $x(t)$ of*

$$\boxed{\dot{x} = \left(\pi_{T_xM} D\hat{F}(x,t)\big|_{T_xM}\right)^{-1} \left(\mathcal{M}(x)F(x,t) + \tfrac{\partial}{\partial t}F(x,t)\right)} \tag{2.31}$$

*exists for all $t \geq 0$ and converges exponentially to $x_*(t)$, provided that $x(0)$ is sufficiently close to $x_*(0)$.*

*Proof.* We show that the conditions $(i)$ and $(ii)$ of Main Theorem 2.1 are satisfied. Due to Lemma 2.4, $(i)$ follows directly from (1). To see $(ii)$, we use the notation of Main Theorem 2.1. Thus, $c_x : [0, R_x] \to M$ denotes a geodesic from $x_*(t)$ to $x \in \mathcal{B}_R(x_*(t))$ with $\|\dot{c}_x(s)\| = 1$ and $R_x := \operatorname{dist}(x_*(t), x) < R$. Moreover, $Y_w$ is a smooth vector field on $\mathcal{B}_R(x_*(t))$ defined by $Y_w(x') := \tau_{x_*(t)x'}\tau^{-1}_{x_*(t)x} w$ for $w \in T_xM$ with $\|w\| = 1$.
Then we get by using Lemma 2.4 for $t \in \mathbb{R}$ and $s \in [0, R_x]$

$$\left\|\nabla_{\dot{c}_x(s)}\nabla_{Y_w}F(c_x(s),t)\right\| = \left\|\nabla_{\dot{c}_x(s)}\pi_{T_{c_x(s)}M} D\hat{F}(c_x(s),t)\cdot Y_w(c_x(s))\right\| =$$

$$\left\|\pi_{T_{c_x(s)}M} D\left(\pi_{T_{c_x(s)}M} D\hat{F}(c_x(s),t)\cdot Y_w(c_x(s))\right)\cdot \dot{c}_x(s)\right\| \leq c_3,$$

due to condition (2) and since $\|Y_w(c_x(s))\| = \|\dot{c}_x(s)\| = 1$.

□

**Time-varying Newton algorithm on submanifolds**
The Euler discretization of the ODE (2.31) computes approximations $x_k$ of $x_*(t_k)$ for $t_k = kh$, $k \in \mathbb{N}$ and $h > 0$. It is given for $\mathcal{M} = -\frac{1}{h}I$ by

$$\boxed{x_{k+1} = \exp_{x_k}\left(-\left(\pi_{T_{x_k}M} D\hat{F}(x_k,t_k)\big|_{T_{x_k}M}\right)^{-1}\left(F(x_k,t_k) + hF^h_\tau(x_k,t_k)\right)\right),} \tag{2.32}$$

where $F^h_\tau(x,t)$ denotes an approximation for $\frac{\partial}{\partial t}F(x,t)$. By applying Main Theorem 2.2 to the situation here, we immediately obtain the following result.

**Theorem 2.4.** *Let $M \subset \mathbb{R}^n$ be a complete Riemannian submanifold with respect to the standard Euclidean inner product on $\mathbb{R}^n$. Let $R > 0$ any real number with $i^*(M) \geq R$. Let $\hat{F} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$, $(x,t) \mapsto \hat{F}(x,t)$ be smooth, let $F = \hat{F}\big|_{M\times\mathbb{R}}$ with $F(x,t) \in T_xM$ and let $t \mapsto x_*(t)$ be a smooth isolated zero of $F$, i.e. $F(x_*(t),t) = 0$ for all $t$. Assume further the existence of constants $c_1, c_2, c_3, c_4, c_5, c_6, c_7 > 0$ such that the following conditions are satisfied for all $t \in \mathbb{R}$*

*(i)* $\|\pi_{T_{x_*(t)}M} D\hat{F}(x_*(t),t)\| \leq c_1$, $\|\frac{\partial}{\partial t}F(x_*(t),t)\| \leq c_2$, $\left\|\pi_{T_{x_*(t)}M} D\hat{F}(x_*(t),t)^{-1}\right\| \leq c_3$,

*(ii)* $\|(\pi_{T_xM}D)^2\hat{F}(x,t)\| \leq c_4$, $\|\frac{\partial^2}{\partial t^2}F(x,t)\| \leq c_5$, $\|\frac{\partial}{\partial t}\pi_{T_xM} D\hat{F}(x,t)\| \leq c_6$ *for* $x \in \mathcal{B}_R(x_*(t))$,

*(iii)* $\|F_\tau^h(x,t) - \frac{\partial}{\partial t}F(x,t)\| \leq c_7 h$, *for all* $x \in \mathcal{B}_R(x_*(t))$, $h > 0$.

*Then for* $c > 0$ *and sufficiently small* $h > 0$, *the sequence defined by* (2.32) *satisfies for* $k \in \mathbb{N}$

$$\operatorname{dist}(x_k, x_*(t_k)) \leq ch,$$

*provided that* $\operatorname{dist}(x_0, x_*(0)) < ch$. *Thus, the update scheme* (2.32) *is well defined and produces estimates for* $x_*(t_k)$, *whose accuracy can be controlled by the step size.*

**Gradient Newton flows**

We now consider the special case where $F$ denotes the intrinsic gradient of a smooth cost function $\Phi$ on a Riemannian submanifold $M \subset \mathbb{R}^n$ endowed with the constant Riemannian metric $I_n$. Thus let

$$\Phi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R},$$

and let $f := \Phi\big|_{M\times\mathbb{R}}$ its restriction to the manifold. Then the intrinsic gradient of $f$ with respect to $x$, $\operatorname{grad} f(x,t) : M \times \mathbb{R} \to TM$, is given by

$$\operatorname{grad} f(x,t) = \pi_{T_xM} \nabla\Phi(x,t),$$

where $\pi_{T_xM} : \mathbb{R}^n \to \mathbb{R}^n$ is the orthogonal projection onto $T_xM$ with kernel $(T_xM)^\perp$. Hence the time-varying Newton flow turns into

$$H_f(x,t) \cdot \dot{x} + \frac{\partial}{\partial t}\operatorname{grad} f(x,t) = \mathcal{M}(x)\operatorname{grad} f(x,t), \tag{2.33}$$

where $H_f(x,t) : T_xM \to T_xM$ denotes the Hessian operator with respect to $x$. It is given by

$$H_f(x,t) = \pi_{T_xM} D\operatorname{grad} f(x,t),$$

cf. Lemma 2.4. Obviously, Theorem 2.4 also holds for $F(x,t) := \operatorname{grad} f(x,t)$ and the update scheme (2.32) turns into

$$\boxed{x_{k+1} = \exp_{x_k}\left(-H_f(x_k,t_k)^{-1}\left(\operatorname{grad} f(x_k,t_k) + hG_h(x_k,t_k)\right)\right),} \tag{2.34}$$

where $G_h$ denotes an approximation of $\frac{\partial}{\partial t}\operatorname{grad} f$.

### 2.1.3.1 Embedding the tracking task in Euclidean space

In order to implement the discrete algorithm (2.32) associated with the time-varying Newton flow on a Riemannian submanifold $M \subset \mathbb{R}^n$, one needs to compute the exponential map and invert the covariant derivative for all $k \in \mathbb{N}$. As these are often very difficult tasks, we now develop techniques, which allow to solve the tracking problem on the manifold by equivalent methods in Euclidean space. This possibly leads to simpler algorithms to track $x_*(t)$ at discrete times $t = t_k$ for $k \in \mathbb{N}$.

**a. Using penalty terms**
Let $M$ be the fiber of a $C^r$-map $g : \mathbb{R}^n \to \mathbb{R}^m$ for $m < n$ such that $M := g^{-1}(0)$. Here, 0 is assumed to be a regular value of $g$, i.e. $\mathrm{rk}Dg(y) = m$ for all $y \in M$ and the dimension of $M$ thus is $(m - n)$.
Let $\Phi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ and let $x_* : \mathbb{R} \to M$ be a continuous minimum of $\Phi$, which additionally satisfies

1. $\Phi(x_*(t), t) = 0$ for all $t \in \mathbb{R}$.
2. There exists $r > 0$ such that for $x \in U_r(x_*(t)) \cap M$ holds: $\Phi(x, t) = 0$ if and only if $x = x_*(t)$.

In order to construct an update scheme to track $x_*(t_k)$ for $k \in \mathbb{N}$, we define a cost function, which augments the constraints into the cost function for $\lambda > 0$ as

$$\hat{\Phi}(x, t) := \Phi(x, t) + \frac{\lambda}{2} g(x)^\top g(x).$$

Then we have

$$\hat{\Phi}(x, t) = 0 \Leftrightarrow \Phi(x, t) = 0 \text{ and } g(x) = 0, \tag{2.35}$$

and therefore, $x_*(t)$ is an isolated minimum of $\hat{\Phi}$. We further define

$$F(x, t) := \nabla\hat{\Phi}(x, t) = \nabla\Phi(x, t) + \frac{\lambda}{2}\nabla(g(x)^\top g(x)),$$

which is a vector valued map in $\mathbb{R}^n$ with smooth isolated zero $x_*(t) \in M$. Hence, we can now use techniques working in Euclidean space to track the zero of $F$ on $M$, cf. Section 2.3.

**b. Lagrangian multipliers**
Here we inspect the classical Lagrangian multiplier technique. Let $M$ be given as the fiber of a smooth map $g : \mathbb{R}^n \to \mathbb{R}^m$, i.e.

$$M = g^{-1}(0),$$

where 0 is a regular value of $g$ and $m < n$. Then $M \subset \mathbb{R}^n$ is a smooth Riemannian submanifold of dimension $n - m$.
Let

$$\Phi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}, \quad (x, t) \mapsto \Phi(x, t)$$

be smooth and let $f$ denote the restriction to $M \times \mathbb{R}$, i.e.

$$f := \Phi\big|_{M\times\mathbb{R}}.$$

Suppose that $f$ has a smooth isolated critical point $x_*(t)$ on $M$, which we want to determine for all $t$. We therefore define

$$L(x, y, t) := \Phi(x, t) + \sum_{i=1}^{m} y_i g_i(x),$$

which is for fixed $t \in \mathbb{R}$ a Lagrangian function and $y_i$ are the Lagrangian multipliers for $i = 1, ..., m$. Thus, the gradient $L$ with respect to $(x, y)$, denoted by $\nabla L$, is given by

$$\nabla L(x, y, t) = \begin{bmatrix} \nabla\Phi(x, t) + \sum_{i=1}^{m} y_i \nabla g_i(x) \\ g(x) \end{bmatrix}.$$

The following result is well known.

**Lemma 2.5.** *Let $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}$ as above. Then $\nabla L(x, y, t) = 0$ if and only if $x \in M$* and $\operatorname{grad}\Phi(x, t) = 0$, *where* $\operatorname{grad}\Phi(x, t) = \pi_{T_xM}\nabla\Phi(x, t)$ *is the intrinsic gradient of $\Phi$.*
*Thus, if $x_*(t)$ is a smooth smooth isolated critical point of $\Phi(x, t)$ on $M$, then there exists a smooth map $y_*(t)$ such that $(x_*(t), y_*(t))$ is a smooth isolated zero of $\nabla L(x, y, t)$.*

*Proof.* "$\Rightarrow$" If $\nabla L(x, y, t) = 0$, then $x \in M$, since $g(x) = 0$. Since $\nabla\Phi(x, t) + \sum_{i=1}^{m} y_i \nabla g_i(x) = 0$, we can conclude that $\nabla\Phi(x, t) \in N_xM$, as $g_i(x) \in N_xM$ for $i = 1, ..., m$. Thus, $x$ is a critical point of the intrinsic gradient of $\Phi\big|_{M\times\mathbb{R}}$
"$\Leftarrow$" Consider for $t \in \mathbb{R}$

$$\nabla L(x_*(t), y, t) = \begin{bmatrix} \nabla\Phi(x_*(t), t) + \sum_{i=1}^{m} y_i \nabla g_i(x_*(t)) \\ g(x_*(t)) \end{bmatrix} =$$

$$\begin{bmatrix} \nabla\Phi(x_*(t), t) + \sum_{i=1}^{m} y_i \nabla g_i(x_*(t)) \\ 0 \end{bmatrix}.$$

Since $x_*(t)$ is a critical point of $\operatorname{grad}\Phi(x, t)$, we have that $\nabla\Phi(x_*(t), t) \in N_{x_*(t)}M$, which implies the existence of $y_*(t) \in \mathbb{R}^m$ such that $\sum_{i=1}^{m} (y_*(t))_i \nabla g_i(x_*(t)) = -\nabla\Phi(x_*(t), t)$, as

$$N_{x_*(t)}M = \operatorname{span}(\nabla g_1(x_*(t)), ..., \nabla g_m(x_*(t))).$$

Since $\nabla g(x_*(t))$ and $\nabla\Phi(x_*(t), t)$ are smooth maps, $y_*(t)$ is also smooth in $t$.

□

This lemma shows, that one can determine a curve of critical points $x_*(t)$ of $\Phi$ on $M$ by tracking a zero of $\nabla L(x, y, t) \in \mathbb{R}^{n+m}$. Thus, we can solve the optimization problem on the manifold by using the tracking algorithms in Euclidean space, as specified in Section 2.3.

**c. Parameterized time-varying Newton algorithm**

We now consider a newer technique, which allows one to pull-back the tracking problem via local coordinate parameterizations from the manifold to an associated tracking problem in Euclidean space. The key feature here is that the coordinate transformations vary with each iteration point. This allows for considerable simplification in the resulting formulas, rather than a fixed set of coordinate charts would do. It enables us also to work on an arbitrary Riemannian manifold. However, for technical reasons we restrict to the simplified situation, where $M \subset \mathbb{R}^n$ is a Riemannian submanifold of Euclidean space. This idea of using local parameterizations has been first used by Shub [61], Hüper and Trumpf [38] and by Manton [47] for time-invariant vector fields and we now extend this approach to our situation.

Thus let $M \subset \mathbb{R}^n$ be a smooth $m$-dimensional Riemannian submanifold endowed with the constant Riemannian metric $I_n$. We consider families of *smooth locally uniform parameterizations* $(\gamma_x)_{x\in P}$ on an open subset $P$ of $M$, i.e. smooth maps

$$\gamma_x : V_x \to U_x \subset M, \qquad \gamma_x(0) = x,$$

such that $\gamma_x$ is a local diffeomorphism around 0. Moreover, we assume that there exists $R > 0$ such that $B_R(0) \subset V_x \subset \mathbb{R}^m$ holds for all $x \in P$. Thus $(\gamma_x)_{x\in P}$ is a system of local coordinate charts around each $x \in P$ that satisfy an uniformity constraint on the sizes of their domains.

For $\tilde{R} > 0$ now choose $P := \{x \in M|\ \operatorname{dist}(x, x_*(t)) < \tilde{R}, t \in \mathbb{R}\}$ and a family $(\gamma_x)_{x\in P}$ of smooth locally uniform parameterizations. Given $x \in P$, let $\hat{\gamma}_x : V_x \times \mathbb{R} \to U_x \times \mathbb{R}$ be defined by $\hat{\gamma}_x(y,t) := (\gamma_x(y), t)$ and consider the pull-back function

$$\Phi \circ \hat{\gamma}_x : V_x \times \mathbb{R} \to \mathbb{R},$$

defined by

$$(y,t) \mapsto \Phi(\gamma_x(y), t).$$

Hence, the $y$-gradient of $\Phi \circ \hat{\gamma}_x$ is

$$\nabla(\Phi \circ \hat{\gamma}_x(y,t)) = \left(\frac{\partial}{\partial y_1}(\Phi \circ \hat{\gamma}_x(y,t)), ..., \frac{\partial}{\partial y_m}(\Phi \circ \hat{\gamma}_x(y,t))\right)^\top$$

and the $y-$Hessian of $\Phi \circ \hat{\gamma}_x$ is given by

$$H_{\Phi\circ\hat{\gamma}_x}(y,t) = \left(\frac{\partial^2}{\partial y_i \partial y_j}(\Phi \circ \hat{\gamma}_x(y,t))\right)_{i,j=1}^m$$

Thus by using two families of smooth local parameterizations $(\gamma_x)_{x\in P}$ and $(\mu_x)_{x\in P}$, we obtain the *parameterized time-varying Newton algorithm*, which is given by

$$\boxed{x_{k+1} = \mu_{x_k}\left(-H_{\Phi\circ\hat{\gamma}_{x_k}}(0,t_k)^{-1}(\nabla(\Phi \circ \hat{\gamma}_{x_k})(0,t_k) + hG^h_{x_k}(0,t_k))\right),} \tag{2.36}$$

where $x_k \in P$ and $G^h_{x_k}(0,t)$ denotes an approximation of $\frac{\partial}{\partial t}\nabla(\Phi \circ \hat{\gamma}_{x_k})(0,t)$, depending on the step size.

In order to get good approximations for the smooth isolated critical points $x_*(t_k) \in M$ of $\Phi(t_k)$, it is necessary, that the two families of parameterizations are quite similar; e.g. equality is allowed. The exact conditions are formulated in the following theorem.

**Main Theorem 2.3.** *Let $M \subset \mathbb{R}^n$ a complete Riemannian submanifold. Let $\Phi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, $(x,t) \mapsto \Phi(x,t)$ be smooth and let $t \mapsto x_*(t)$ be a smooth isolated critical point of $\Phi$ on $M$. Let for some $R, \tilde{R} > 0$, $\gamma_x : V_x \to U_x \subset M$ and $\mu_x : V_x \to U'_x \subset M$ denote for $x \in P$ families of local parameterizations such that $B_R(0) \subset V_x$, where $P := \{x \in M|\ \operatorname{dist}(x, x_*(t)) < \tilde{R}, t \in \mathbb{R}\}$. Assume further the existence of constants $c_1, c_2, c_3, c_4, c_5, c_6, c_7, m_1, m_2,\ m_3 > 0$ such that* *(i)* $\|H_{\Phi\circ\hat{\gamma}_{x_*(t)}}(0,t)\| \leq c_1$ *for all* $t \in \mathbb{R}$,

*(ii)* $\|\frac{\partial}{\partial t}\nabla(\Phi \circ \hat{\gamma}_x)(0,t)\| \leq c_2$, $\|H_{\Phi\circ\hat{\gamma}_x}(0,t)^{-1}\| \leq c_3$ *for all* $x \in \mathcal{B}_{\tilde{R}}(x_*(t))$, $t \in \mathbb{R}$,

*(iii)* $\|\frac{\partial}{\partial y}H_{\Phi\circ\hat{\gamma}_x}(y,t)\| \leq c_4$, $\|\frac{\partial^2}{\partial t^2}\nabla(\Phi \circ \hat{\gamma}_x)(y,t)\| \leq c_5$, $\|\frac{\partial}{\partial t}H_{\Phi\circ\hat{\gamma}_x}(0,t)\| \leq c_6$ *for all* $x \in \mathcal{B}_{\tilde{R}}(x_*(t))$ *and* $y \in B_R(0)$, $t \in \mathbb{R}$,

*(iv)* $(G_x^h)_{x\in P}$ *denotes a family of maps such that for* $t \in \mathbb{R}$, $h > 0$ *and* $x \in \mathcal{B}_{\tilde{R}}(x_*(t))$

$$\left\|G_x^h(0,t) - \frac{\partial}{\partial t}\nabla(\Phi \circ \hat{\gamma}_x)(0,t)\right\| \leq c_7 h,$$

*(v)* $\sigma_{min}\left(D\gamma_x(0)\right) \geq m_1$, $\sigma_{max}\left(D\gamma_x(0)\right) \leq m_2$, $\|D^2\gamma_x(y)\| \leq m_3$ *for all* $x \in \mathcal{B}_{\tilde{R}}(x_*(t))$ *and* $y \in B_R(0)$*, where* $\sigma_{min}$ *(*$\sigma_{max}$*) denotes the smallest (largest) singular value.*

*(vi)* $\mu_x(0) = \gamma_x(0)$, $D\mu_x(0) = D\gamma_x(0)$, $\|D^2\mu_x(y)\| \leq m_3$ *for all* $x \in \mathcal{B}_{\tilde{R}}(x_*(t))$ *and* $y \in B_R(0)$.

*Then the following statements hold for* $R' = \min\{R, \frac{m_1}{4m_3}, \frac{1}{2c_3c_4}, \frac{c_3}{2c_4}\}$ *and for* $\hat{R} = \min\{\frac{3m_1}{4}R', \tilde{R}\}$

1. *For any* $x \in P$*, the parameterization* $\gamma_x$ *satisfies for all* $y \in B_{R'}(0)$

$$\sigma_{min}\left(D\gamma_x(y)\right) \geq \frac{3}{4}m_1.$$

   *This moreover shows that* $\mathcal{B}_r(x) \subseteq \gamma_x(B_{R'}(0))$ *for* $r = \frac{3}{4}m_1R'$*. Thus we have* $x_*(t) \in \gamma_x(B_{R'}(0))$ *for* $x \in \mathcal{B}_{\hat{R}}(x_*(t))$*.*

   *Note that the same claims hold for the parameterization* $\mu_x$*, since* $\mu_x$ *satisfies the same conditions as* $\gamma_x$*.*

2. *The following equation holds for all* $t \in \mathbb{R}$, $x \in \mathcal{B}_{\hat{R}}(x_*(t))$ *and* $y \in B_{R'}(0)$

$$\frac{m_1}{4}\|\gamma_x^{-1}(x_*(t)) - y\| \leq \operatorname{dist}(\gamma_x(y), x_*(t)) \leq (m_2 + 2m_3R')\|\gamma_x^{-1}(x_*(t)) - y\|.$$

3. *There exists* $c_8, c_9 > 0$ *such that for* $t \in \mathbb{R}$, $x \in \mathcal{B}_{\hat{R}}(x_*(t))$ *and* $y \in B_{R'}(0)$

$$\operatorname{dist}(\gamma_x(y), x_*(t)) \leq c_8\|\nabla(\Phi \circ \hat{\gamma}_x)(y,t)\|, \tag{2.37}$$

   *and*

$$\|\nabla(\Phi \circ \hat{\gamma}_x)(y,t)\| \leq c_9\operatorname{dist}(\gamma_x(y), x_*(t)). \tag{2.38}$$

4. *The discretization sequence* $(x_k)$ *as defined in* (2.36) *with* $t_k = kh$, $h > 0$ *satisfies for some* $c_{10}, c_{11} > 0$

$$\operatorname{dist}(x_{k+1}, x_*(t_{k+1})) \leq c_{10}\operatorname{dist}(x_k, x_*(t_k))^2 + c_{11}h^2 \tag{2.39}$$

   *for* $x_*(t_k), x_*(t_k + h), x_{k+1} \in \mathcal{B}_{\hat{R}}(x_k)$, $k \in \mathbb{N}_0$.

5. *Let* $c > 0$ *be constant and* $h$ *sufficiently small. For any initial condition* $x_0$ *with* $\operatorname{dist}(x_0, x_*(0)) < ch$ *we have*

$$\operatorname{dist}(x_k, x_*(t_k)) \leq ch$$

*for all* $k \in \mathbb{N}_0$*. Thus, the update scheme* (2.36) *is well defined and produces estimates for* $x_*(t_k)$*, whose accuracy can be controlled by the step size.*

*Proof.* 1) We have for $\|y\| \leq R' \leq \frac{m_1}{4m_3}$

$$D\gamma_x(y) = D\gamma_x(0) + \mathcal{R},$$

where $\|\mathcal{R}\| \leq \|y\| m_3$. Thus

$$\sigma_{min}\left(D\gamma_x(y)\right) \geq \sigma_{min}\left(D\gamma_x(0)\right) - m_3 R' \geq m_1 - m_3 \frac{m_1}{4m_3} = \frac{3}{4} m_1.$$

To prove the second claim, consider

$$\gamma_x(y) = \gamma_x(0) + D\gamma_x(0)y + \mathcal{R},$$

where $\|\mathcal{R}\| \leq m_3 \|y\|^2$. Thus,

$$\gamma_x(y) - \gamma_x(0) = D\gamma_x(0)y + \mathcal{R},$$

which implies that

$$\|\gamma_x(y) - \gamma_x(0)\| \geq \frac{3m_1}{4}\|y\|,$$

since $\|y\| \leq \frac{m_1}{4m_3}$. This moreover shows, that

$$\operatorname{dist}(\gamma_x(y), \gamma_x(0)) \geq \frac{3m_1}{4}\|y\|. \tag{2.40}$$

Now we consider for $\epsilon > 0$ the border $\partial B_{R'-\epsilon}(0) := \{y \in \mathbb{R}^m | \ \|y\| = R' - \epsilon\}$ of $B_{R'-\epsilon}(0)$. Since $\gamma_x$ is a local parameterization of $M$, $\gamma_x(\partial B_{R'-\epsilon}(0))$ is the border of $\gamma_x(B_{R'-\epsilon}(0))$ in $M$. Due to the above equation, we have for $\bar{x} \in \partial\gamma_x(B_{R'-\epsilon}(0))$ and $\bar{y} := \gamma_x^{-1}(\bar{x}) \in \partial B_{R'-\epsilon}(0)$

$$\operatorname{dist}(x, \bar{x}) = \operatorname{dist}(\gamma_x(0), \gamma_x(\bar{y})) \geq \frac{3m_1}{4}\|\bar{y}\| = \frac{3m_1}{4}(R' - \epsilon),$$

which shows that $\mathcal{B}_{\frac{3m_1}{4}(R'-\epsilon)}(x) \subseteq \gamma_x(B_{R'-\epsilon}(0))$. Since this relation holds for all $\epsilon > 0$, the claim follows.

2) To show the right inequality, let $\alpha : [0,1] \to V_x$, $s \mapsto y + (y_* - y)s$ be a curve from $y$ to $y_* := \gamma_x^{-1}(x_*(t))$. Thus

$$\operatorname{dist}(\gamma_x(y), x_*(t)) \leq \int_0^1 \|\frac{d}{ds}(\gamma_x \circ \alpha)(s))\| ds = \int_0^1 \|D\gamma_x(\alpha(s))(y_* - y)\| ds \leq$$

$$\leq \max_{s\in[0,1]} \|D\gamma_x(\alpha(s))\|\|y_* - y\| \leq (m_2 + m_3\|y_* - y\|)\|y_* - y\| \leq (m_2 + 2m_3R')\|y_* - y\|,$$

since $\|y\| \leq R'$ and $\|y_*\| \leq R'$.
The left inequality of the claimed estimate can be seen by considering

$$\gamma_x(y_*) = \gamma_x(y) + D\gamma_x(y)(y_* - y) + \mathcal{R},$$

where $\|\mathcal{R}\| \leq m_3\|y_* - y\|^2$. Thus,

$$\gamma_x(y_*) - \gamma_x(y) = D\gamma_x(y)(y_* - y) + \mathcal{R},$$

which implies that

$$\|\gamma_x(y_*) - \gamma_x(y)\| \geq \frac{m_1}{4}\|y_* - y\|,$$

since $\|y_* - y\| \leq \frac{2m_1}{4m_3}$. This moreover shows, that

$$\mathrm{dist}(\gamma_x(y_*), \gamma_x(y)) \geq \frac{m_1}{4}\|y_* - y\|. \tag{2.41}$$

3) Let $y_* = \gamma_x^{-1}(x_*(t))$ and consider the Taylor series

$$\nabla(\Phi \circ \hat{\gamma}_x)(y_*, t) = \nabla(\Phi \circ \hat{\gamma}_x)(y, t) + H_{\Phi\circ\hat{\gamma}_x}(y, t)(y_* - y) + \mathcal{R} \tag{2.42}$$

where $\mathcal{R}$ satisfies:

$$\|\mathcal{R}\| \leq c_4\|y_* - y\|^2.$$

Since $\nabla(\Phi \circ \hat{\gamma}_x)(y_*, t) = 0$, equation (2.42) is equivalent to

$$\nabla(\Phi \circ \hat{\gamma}_x)(y, t) + H_{\Phi\circ\hat{\gamma}_x}(y, t)(y_* - y) + \mathcal{R} = 0.$$

We therefore get

$$\|y_* - y\| \leq \left\|H_{\Phi\circ\hat{\gamma}_x}(y, t)^{-1}(\nabla(\Phi \circ \hat{\gamma}_x)(y, t) + \mathcal{R})\right\|,$$

implying that

$$\|y_* - y\| \leq \frac{c_3}{2}(\|\nabla(\Phi \circ \hat{\gamma}_x)(y, t)\| + c_4\|y_* - y\|^2),$$

since $\|H_{\Phi\circ\hat{\gamma}_x}(y, t)^{-1}(\nabla(\Phi \circ \hat{\gamma}_x)(y, t)\| \leq \frac{c_3}{2}$ for $\|y\| \leq \frac{c_3}{2c_4}$ (Taylor).
Hence the above equation is equivalent to

$$\|y_* - y\| \left(1 - \frac{c_3c_4}{2}\|y_* - y\|\right) \leq \frac{c_3}{2}\|\nabla(\Phi \circ \hat{\gamma}_x)(y, t)\|.$$

Hence, if $\|y_* - y\| \leq \frac{1}{c_3c_4}$, then

$$\|y_* - y\| \leq c_3\|\nabla(\Phi \circ \hat{\gamma}_x)(y, t)\|.$$

By taking claim 2) into account, (2.37) gets obvious.

To see (2.38), we also use the second claim and note that equation (2.42) implies that

$$\|\nabla(\Phi \circ \hat{\gamma}_x)(y,t)\| \le \|H_{\Phi\circ\hat{\gamma}_x}(y,t)(y_* - y)\| + \|\mathcal{R}\| \le \|y_* - y\|(c_1 + c_4\|y_* - y\|).$$

4) Let $x_k = \gamma_{x_k}(0)$ and $x_{k+1} = \mu_{x_k}(y)$ for some $y \in \mathbb{R}^m$. Consider at first the Taylor series:

$$\nabla(\Phi\circ\hat{\gamma}_{x_k})(y,t_{k+1}) = \nabla(\Phi\circ\hat{\gamma}_{x_k})(0,t_k) + \frac{\partial}{\partial t}\nabla(\Phi\circ\hat{\gamma}_{x_k})(0,t_k)h + H_{\Phi\circ\hat{\gamma}_{x_k}}(0,t_k)y + \mathcal{R}, \quad (2.43)$$

where $\mathcal{R}$ satisfies:

$$\|\mathcal{R}\| \le c_5h^2 + c_4\|y\|^2 + c_6h\|y\| \le \left(c_4 + \frac{c_6}{2}\right)\|y\|^2 + \left(c_5 + \frac{c_6}{2}\right)h^2$$

The update scheme (2.36) requires that

$$y = -H_{\Phi\circ\hat{\gamma}_{x_k}}(0,t_k)^{-1}(\nabla(\Phi\circ\hat{\gamma}_{x_k})(0,t_k) + hG^h_{x_k}(0,t_k)).$$

Therefore, equation (2.43) simplifies to

$$\nabla(\Phi\circ\hat{\gamma}_{x_k})(y,t_{k+1}) = \frac{\partial}{\partial t}\nabla(\Phi\circ\hat{\gamma}_{x_k})(0,t_k)h - hG^h_{x_k}(0,t_k) + \mathcal{R}, \quad (2.44)$$

and thus

$$\|\nabla(\Phi\circ\hat{\gamma}_{x_k})(y,t_{k+1})\| \le c_7h^2 + \left(c_4 + \frac{c_6}{2}\right)\|y\|^2 + \left(c_5 + \frac{c_6}{2}\right)h^2. \quad (2.45)$$

Now consider

$$\|y\| = \left\|H_{\Phi\circ\hat{\gamma}_{x_k}}(0,t_k)^{-1}(\nabla(\Phi\circ\hat{\gamma}_{x_k})(0,t_k) + hG^h_{x_k}(0,t_k))\right\|,$$

which shows that

$$\|y\| \le c_3(\|\nabla(\Phi\circ\hat{\gamma}_{x_k})(0,t_k)\| + h\|G^h_{x_k}(0,t_k)\|). \quad (2.46)$$

Note that $\|G^h_{x_k}(0,t_k)\| \le \|\frac{\partial}{\partial t}\nabla(\Phi\circ\hat{\gamma}_{x_k})(0,t_k)\| + c_7h \le c_2 + c_7h$. Then (2.46) turns into

$$\|y\| \le c_3\big(\|\nabla(\Phi\circ\hat{\gamma}_{x_k})(0,t_k)\| + hc_2 + h^2c_7\big),$$

which implies that

$$\|y\| \le c_3\big(c_9\mathrm{dist}(x_k, x_*(t_k)) + h(c_2 + c_7h)\big).$$

By using the abbreviations $k_1 = c_3c_9$ and $k_2 = c_3(c_2 + c_7h)$, this equation implies that

$$\|y\|^2 \le k_1^2\mathrm{dist}(x_k, x_*(t_k))^2 + k_2^2h^2 + 2k_1k_2\mathrm{dist}(x_k, x_*(t_k))h,$$

hence,

$$\|y\|^2 \le (k_1^2 + k_1k_2)\mathrm{dist}(x_k, x_*(t_k))^2 + (k_2^2 + k_1k_2)h^2.$$

Plug this into (2.45) and obtain

$$\|\nabla(\Phi \circ \hat{\gamma}_{x_k})(y, t_{k+1})\| \leq$$

$$k_3(k_1^2 + k_1k_2)\mathrm{dist}(x_k, x_*(t_k))^2 + \left(k_3k_2^2 + k_1k_2k_3 + c_7 + c_5 + \frac{c_6}{2}\right)h^2,$$

where $k_3 = (c_4 + c_6/2)$. By using (2.37) we get

$$\mathrm{dist}(\gamma_{x_k}(y), x_*(t_{k+1})) \leq$$

$$c_8k_3(k_1^2 + k_1k_2)\mathrm{dist}(x_k, x_*(t_k))^2 + c_8(k_3k_2^2 + k_1k_2k_3 + c_7 + c_5 + \frac{c_6}{2})h^2.$$

From this, we conclude that

$$\mathrm{dist}(x_{k+1}, x_*(t_{k+1})) = \mathrm{dist}(\mu_{x_k}(y), x_*(t_{k+1})) \leq$$

$$c_8k_3(k_1^2 + k_1k_2)\mathrm{dist}(x_k, x_*(t_k))^2 + (c_8(k_3k_2^2 + k_1k_2k_3 + c_7 + c_5 + \frac{c_6}{2})h^2 +$$

$$m_3((k_1^2 + k_1k_2)\mathrm{dist}(x_k, x_*(t_k))^2 + (k_2^2 + k_1k_2)h^2),$$

since $\mu_{x_k}(y) = \gamma_{x_k}(y) + \mathcal{R}$, where $\|\mathcal{R}\| \leq m_3\|y\|^2$. This shows (4).

5) Note that for sufficiently small $h > 0$ holds that $x_*(t_k + h), x_{k+1} \in \mathcal{B}_{\hat{R}}(x_k)$ for $x_*(t_k) \in \mathcal{B}_{\hat{R}/2}(x_k)$. Thus let for sufficiently small $h > 0$ and $ch \leq \hat{R}/2$,

$$\mathrm{dist}(x_k, x_*(t_k)) \leq ch,$$

for some $k$. By the estimate (2.39) then

$$\mathrm{dist}(x_{k+1}, x_*(t_{k+1})) \leq$$

$$c_{10}\mathrm{dist}(x_k, x_*(t_k))^2 + c_{11}h^2 \leq c_{10}c^2h^2 + c_{11}h^2 \leq ch, \tag{2.47}$$

for $h < \frac{c}{c_{10}c^2 + c_{11}}$.

□

**Remark 2.4.**

1. Note that valid choices for approximations $G_x^h(0,t)$ of $\frac{\partial}{\partial t}\nabla(\Phi \circ \hat{\gamma}_x)(0,t)$ are given in Section 2.1.2.2. For example,

   $$G_x^h(0,t) := \frac{1}{h}\left(\nabla(\Phi \circ \hat{\gamma}_x)(0,t) - \nabla(\Phi \circ \hat{\gamma}_x)(0,t-h)\right)$$

   is a 1st order approximation of $\frac{\partial}{\partial t}\nabla(\Phi \circ \hat{\gamma}_x)(0,t)$.

2. On compact manifolds such parameterizations $\gamma_x$, $x \in M$, always exist, since the exponential maps can be used to construct them; e.g. $\gamma_x := \exp_x \circ \tau_x$, where $\tau_x : \mathbb{R}^m \to T_xM$ is a linear map.

## 2.2 Newton flows on Lie groups

In the case of working on a Lie group $G$ one is tempted to consider Riemannian metrics that are linked to the Lie group structure. These are the left-, right- or bi-invariant Riemannian metrics. Of course, it is also possible to consider the Newton flow for an arbitrary Riemannian metric on $G$. However, for invariant Riemannian metrics, there are explicit expressions for the Levi-Civita connection that simplify the computations of the Newton flow. We illustrate such possibilities by the examples below. Thus let $G$ denote a Lie group of dimension $n$, endowed with a left invariant metric $\langle , \rangle_g$ and let $e_1, ..., e_n$ denote a basis of the Lie algebra $\mathfrak{g}$.
Let further $l_g : G \to G$, $h \mapsto gh$, denote the left translation by an element $g \in G$ and let $L_g : \mathfrak{g} \to T_gG$ be the linear isomorphism defined by the tangent map at $e$ of the left translation, i.e.

$$L_g(X) = T_e l_g(X).$$

Then one has a frame $E_1, ..., E_n$ of left invariant vector fields on $G$ by setting

$$E_i(g) = T_e l_g \cdot e_i, \quad i = 1, ..., k.$$

A metric $\langle , \rangle_g$ is said to be left invariant, if for all $g \in M$ and $u, v \in T_gM$ holds

$$\langle u, v\rangle_g = \langle T_g l_{g^{-1}} \cdot u, T_g l_{g^{-1}} \cdot v\rangle_e.$$

Analogously, one can define right invariant metrics by using the right translation $r_g$, which is needed to introduce bi-invariant metrics.

In order to derive explicit formulas, we have to study the relation between connections on Lie groups and bilinear maps

$$\omega : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}.$$

According to [34], $\omega$ uniquely defines a left invariant connection $\nabla$ on $G$ by demanding that

$$(\nabla_{E_i} E_j)(e) = \omega(e_i, e_j)$$

for all $i, j = 1, ..., n$. Then $\omega$ is called the connection form of $\nabla$ and for arbitrary vector fields $X = \sum_{i=1}^n \phi_i E_i, Y = \sum_{i=1}^n \psi_i E_i$, the connection is given as

$$\nabla_X Y(g) = L_g \left( \omega \left( \sum_{i=1}^n \phi_i(g) e_i, \sum_{j=1}^n \psi_j(g) e_j \right) + \sum_{j=1}^n d\psi_j(g) X(g) e_j \right).$$

Thus, by choosing special connection forms $\omega$, one obtains different types of connections. For the Levi-Civita connection, in particular, one has the following result ([34]).

**Theorem 2.5.** *Let $G$ be a Lie group, endowed with a left invariant Riemannian metric induced by an inner product $\langle , \rangle_{\mathfrak{g}}$ on the Lie algebra $\mathfrak{g}$.*

(a) *Then the Levi-Civita connection is the unique left invariant connection with connection form*

$$\omega(x,y) = \frac{1}{2}\left([x,y] - (\operatorname{ad} x)^* y - (\operatorname{ad} y)^* x\right).$$

(b) *If $\langle , \rangle_g$ defines a bi-invariant Riemannian metric on $G$, then $(\operatorname{ad} x)^* = -\operatorname{ad} x$ and therefore the Levi-Civita connection has the connection form*

$$\omega(x,y) = \frac{1}{2}[x,y].$$

Here, $(\operatorname{ad} x)^* : \mathfrak{g} \to \mathfrak{g}$ denotes the adjoint operator to $\operatorname{ad} x$ with respect to the inner product $\langle , \rangle_g$. Thus $(\operatorname{ad} x)^*$ is uniquely defined by

$$\langle (\operatorname{ad} x)^* y, z\rangle_g = \langle y, (\operatorname{ad} x) z\rangle_g = \langle y, [x,z]\rangle_g.$$

The previous theorem shows that in the case of a bi-invariant metric $\langle , \rangle_g$, the Levi-Civita connection $\nabla$ is given by

$$\nabla_X Y(g) = L_g \left( \frac{1}{2} \left[ \sum_{i=1}^{n} \phi_i(g) e_i, \sum_{j=1}^{n} \psi_j(g) e_j \right] + \sum_{j=1}^{n} d\psi_j(g) X(g) e_j \right). \tag{2.48}$$

Note that there always exist bi-invariant metrics on compact Lie groups, cf. Proposition 4.24 in [42].

**Proposition 2.4.** *Let $G$ denote a compact Lie group, endowed with a bi-invariant metric $\langle , \rangle_g$. Then for $L_g = g$, $\dot{g} = g\Omega(g) \in T_g G$ and $F(g) = g\tilde{\Omega}(g) \in T_g G$ the Levi-Civita connection is given by*

$$\nabla_{\dot{g}} F(g) = g \left( \frac{1}{2}[\Omega, \tilde{\Omega}] + D\tilde{\Omega}(g) \cdot \dot{g} \right). \tag{2.49}$$

*Proof.* The claimed formula follows directly from (2.48) by noting that

$$\dot{g} = \sum_{i=1}^{n} \phi_i(g) E_i = \sum_{i=1}^{n} \phi_i(g) g e_i = g \sum_{i=1}^{n} \phi_i(g) e_i := g\Omega(g),$$

$$F(g) = \sum_{j=1}^{n} \psi_j(g) E_j = g \sum_{j=1}^{n} \psi_j(g) e_j = g\tilde{\Omega},$$

and

$$\sum_{j=1}^{n} d\psi_j(g) X(g) e_j = \sum_{j=1}^{n} d\psi_j(g) \dot{g}(g) e_j = D\tilde{\Omega} \cdot \dot{g}.$$

□

From now on we assume that $G$ is a Lie group that is endowed with a bi–invariant Riemannian metric. In order to simplify our notation we write $g\xi$ for the left translation $L_g(\xi)$ and $g\xi g^{-1}$ for the adjoint action $Ad(g)\xi$ on $\mathfrak{g}$. Let $\exp : \mathfrak{g} \to G, \exp(\xi) = e^{\xi}$ denote the exponential map on the Lie algebra $\mathfrak{g}$. Since the connection form $\omega : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ for the Levi-Civita connection satisfies $\omega(\xi, \xi) = \frac{1}{2}[\xi, \xi] = 0$ for all $\xi \in \mathfrak{g}$, we see that the exponential curves $c : \mathbb{R} \to G$

$$c(t) = ge^{t\xi}$$

are geodesics. Therefore, the Riemannian exponential map $\exp_g : T_gG \to G$ is given as

$$\exp_g(g\xi) = ge^{\xi}$$

for all $\xi \in \mathfrak{g}$. We prove

**Proposition 2.5.** *Let $G$ denote a connected Lie group, endowed with a bi-invariant Riemannian metric and let* $\exp : \mathfrak{g} \to G$ *denote the exponential map. Then*

*(i) $G$ is geodesically complete and the geodesics are $c : \mathbb{R} \to G, c(t) = ge^{t\xi}$ for all $g \in G$, $\xi \in \mathfrak{g}$. The Riemannian exponential map is*

$$\exp_g : T_gG \to G, \exp_g(\zeta) = g \exp(g^{-1}\zeta), \quad \forall \zeta \in T_gG.$$

*Moreover,* $\exp_g$ *is surjective for any $g \in G$.*

*(ii) Let $c : [0, 1] \to G$ denote the unique geodesic with $c(0) = g, c(1) = h$. Thus $c(t) = ge^{t\xi}$ with $e^{\xi} = g^{-1}h$. Then the unique parallel vector field $V : [0, 1] \to TG$, $V(t) \in T_{c(t)}G$, along $c$ with $V(0) = g\eta$ is given as*

$$V(t) = c(t)e^{-t\xi/2}\eta e^{t\xi/2}.$$

*In particular, the parallel translation along $c$ from $g$ to $h$ in $G$ is given as $\tau_{gh} : T_gG \to T_hG$,*

$$\tau_{gh}(g\eta) = he^{-\xi/2}\eta e^{\xi/2}, \quad \xi := g^{-1}\dot{c}(0).$$

*Proof.* For (i) note that the formulas for the geodesics and Riemannian exponential map are already shown. Since the exponential map $ge^{t\xi}$ exists for all $t \in \mathbb{R}$ it follows that $G$ is geodesically complete. The Hopf-Rinow theorem then implies that any two points in $G$ can be joined by a geodesic. This shows the surjectivity of $\exp_g$. For (ii), we have to show that the covariant derivative of $V$ along $c$ is identically zero. In fact, for $\Omega = e^{-t\xi/2}\eta e^{t\xi/2}$ we have $\dot{\Omega} = -\frac{1}{2}[\xi, \Omega]$ and therefore

$$\nabla_{\dot{c}}V = c(t)\left(\frac{1}{2}[\xi, \Omega] + \dot{\Omega}\right) = 0.$$

This completes the proof. □

We now turn to the task of comparing the covariant derivative of a vector field along a geodesic at two different points. Thus let $X : G \to TG$ denote a smooth vector field. By left translation this yields a smooth map

$$\Omega : G \to \mathfrak{g}, \quad \Omega(g) := g^{-1}X(g).$$

In the sequel we need to compute the second derivative of $\Omega$. This is defined as follows.

**Definition 2.4.** *The second derivative of a smooth map $\Omega : G \to \mathfrak{g}$ at $g \in G$ is the symmetric bilinear map*

$$D^2\Omega(g) : T_gG \times T_gG \to \mathfrak{g}$$

*defined by polarization from the quadratic form*

$$D^2\Omega(g)(g\xi, g\xi) := \frac{d^2}{dt^2}\Omega(c(t)),$$

*where $c$ is the geodesic $c(t) = ge^{t\xi}$.*

**Theorem 2.6.** *Let $G$ be a compact Lie group, endowed with the bi-invariant Riemannian metric defined by the Killing form. Let $R$ denote the injectivity radius of $G$ and $0 < r < R$. For any two elements $g, h \in G$ with distance $\operatorname{dist}(g, h) \leq r$ and any tangent vector $v \in T_gG$ let $w = \tau_{gh}v$ denote the parallel transport along the unique geodesic $c : [0, 1] \to G$ connecting $g$ with $h$. Let $X$ be a smooth vector field on $G$ such that $\Omega : G \to \mathfrak{g}, \Omega(g) := g^{-1}X(g)$ satisfies $||D^2\Omega(c(t))|| \leq \gamma$, for some constant $\gamma > 0$ and all $t \in [0, 1]$. Then*

$$||\tau_{gh}\nabla_v X(g) - \nabla_w X(h)|| \leq C\operatorname{dist}(g, h)||v||$$

*holds for*

$$C := 3\max||X(c(t))|| + 2||D\Omega(c(0))|| + 3\gamma.$$

*Proof.* Let $c : [0, 1] \to G, c(t) = g\exp(t\xi)$ denote the unique geodesic connecting $g$ with $h$. Thus $\operatorname{dist}(g, h) = ||\xi||$. Let $V : [0, 1] \to TG$ denote the unique vector field along $c$ obtained by parallel translation of the vector $v$ along $c$. Thus

$$V(t) = c(t)e^{-t\xi/2}ve^{t\xi/2}.$$

Let $\Omega(g) := g^{-1}X(g)$ and $\tilde{\Omega}(t) := e^{-t\xi/2}ve^{t\xi/2}$. Then

$$\nabla_V X = c(t)\left(\frac{1}{2}[\tilde{\Omega}, \Omega] + D\tilde{\Omega}(c(t))V(t)\right)$$

and thus

$$\nabla_{\dot{c}}\nabla_V X =$$

$$c(t)\left(\frac{1}{4}[\xi, [\tilde{\Omega}, \Omega]] + \frac{1}{2}[\xi, D\Omega(c(t))V(t)] + \frac{1}{2}[\dot{\tilde{\Omega}}, \Omega] + \frac{1}{2}[\tilde{\Omega}, \dot{\Omega}] + D^2\Omega(c(t))(c(t)\xi, V(t))\right).$$

Note that, for the Killing form, we have for Lie algebra elements $x, y$ the estimate

$$||[x, y]|| \leq 2||x||||y||.$$

Note further, that $||\tilde{\Omega}|| = ||v||$.
Therefore

$$||\nabla_{\dot{c}}\nabla_V X|| \leq ||\xi||||v|| \left(3||\Omega(c(t))|| + 2||D\Omega(c(t))|| + ||D^2\Omega(c(t))||\right). \tag{2.50}$$

By assumption,

$$||D^2\Omega(c(t))|| \leq \gamma$$

and therefore also

$$||D\Omega(c(t))|| \leq ||D\Omega(c(0))|| + \gamma.$$

The result follows. □

**Time-varying Newton flow on Lie groups**
We consider a time-varying vector field $F : G \times \mathbb{R} \to TG$, $(g, t) \mapsto F(g, t)$. This vector field can be rewritten as

$$F(g, t) = \sum_{i=1}^{n} \psi_i(g, t)E_i,$$

by using suitable functions $\psi_i(g, t)$, cf. above.
We assume, that there exists a continuous map $g_* : \mathbb{R} \to G$ such that for all $t \in \mathbb{R}$ holds

$$F(g_*(t), t) = 0.$$

We want to use the time-varying Newton flow, in order to track the zero $g_*(t)$ of $F$. Thus let further $\dot{g}(g) = \sum_{i=1}^{n} \phi_i(g, t)E_i$. Then the Levi-Civita connection is given by

$$\nabla_{\dot{g}}F(g, t) = L_g\left(\omega\left(\sum_{i=1}^{n} \phi_i(g, t)e_i, \sum_{j=1}^{n} \psi_j(g, t)e_j\right) + \sum_{j=1}^{n} d\psi_j(g, t)\dot{g}(g)e_j\right),$$

and the time-varying Newton flow (2.12) turns into

$$L_g\left(\omega\left(\sum_{i=1}^{n} \phi_i(g, t)e_i, \sum_{j=1}^{n} \psi_j(g, t)e_j\right) + \sum_{j=1}^{n} d\psi_j(g, t)X(g)e_j\right) + \frac{\partial}{\partial t}F(g, t)$$
$$= \mathcal{M}(g)F(g, t),$$

where $\mathcal{M}$ is a stable bundle map. In the case of a bi-invariant metric we have $L_g = g$, $\dot{g} = g\Omega(g, t)$ and $F(g, t) = g\tilde{\Omega}(g, t)$. The Newton flow turns into

$$g\left(\frac{1}{2}[\Omega(g, t), \tilde{\Omega}(g, t)] + D\tilde{\Omega}(g, t) \cdot \dot{g}\right) + \frac{\partial}{\partial t}F(g, t) = \mathcal{M}(g)F(g, t). \tag{2.51}$$

This is equivalent to

$$\frac{1}{2}[\Omega(g, t), \tilde{\Omega}(g, t)] + D\tilde{\Omega}(g, t) \cdot \dot{g} = g^{-1}\mathcal{M}(g)g\tilde{\Omega}(g, t) - \frac{\partial}{\partial t}\tilde{\Omega}(g, t),$$

which shows, that an explicit expression of (2.51) is given by

$$\boxed{\dot{g} = g\Omega(g,t),} \tag{2.52}$$

where $\Omega(g,t)$ satisfies

$$\frac{1}{2}[\Omega(g,t), \tilde{\Omega}(g,t)] + D\tilde{\Omega}(g,t) \cdot g\Omega(g,t) = \mathcal{M}\tilde{\Omega}(g,t) - \frac{\partial}{\partial t}\tilde{\Omega}(g,t).$$

By reformulating Main Theorem 2.1 for this special case, we get at the following

**Theorem 2.7.** *Let $G$ denote a Lie group endowed with a bi-invariant metric and let $R > 0$ be any real number with $i^*(G) \geq R$ . Let $F : G \times \mathbb{R} \to TG$, $(g,t) \mapsto F(g,t) =: g\tilde{\Omega}(g,t)$ be a smooth vector field on $G$ and let $t \mapsto g_*(t)$ denote a smooth isolated zero of $F$ on $G$, i.e. $F(g_*(t),t) = 0$. Assume further the existence of constants $c_1, c_2, c_3 > 0$ such that the following conditions are satisfied for $t \in \mathbb{R}$*

1. *$c_1 \leq \|\nabla_v F(g_*(t),t)\| \leq c_2$ for all $v \in T_{g_*(t)}G$ with $\|v\| = 1$,*
2. *$\|\nabla_{\dot{c}_g(s)} \nabla_{Y_w} F(c_g(s),t)\| \leq c_3$ for all $g \in \mathcal{B}_R(g_*(t))$, $s \in [0, R_g]$ and $w \in T_gG$ with $\|w\| = 1$.*

   *Here $R_g := \operatorname{dist}(g_*(t), g)$ and $c_g : [0, R_g] \to G$ is a geodesic from $g_*(t)$ to $g$ with $\|\dot{c}_g(s)\| = 1$ for $s \in [0, R_g]$ and $Y_w$ is a smooth vector field on $\mathcal{B}_R(g_*(t))$ defined via the parallel transport by $Y_w(g') := \tau_{g_*(t)g'}\tau^{-1}_{g_*(t)g}w$.*

*Then for $g(0)$ sufficiently close to $g_*(0)$, the solution $g(t)$ of (2.52) exists for all $t \geq 0$ and converges exponentially to $g_*(t)$.*

**Remark 2.5.** In the case of working on a compact Lie group endowed with the bi-invariant metric defined by the killing form, we get from (2.50), that the following estimate holds

$$||\nabla_{\dot{c}} \nabla_V F(c(s),t)|| \leq ||\xi|| ||v|| \left(3||\tilde{\Omega}(c(s))|| + 2||D\tilde{\Omega}(c(s))|| + ||D^2\tilde{\Omega}(c(s))||\right).$$

Here, $c : [0,1] \to G$, defined by $s \mapsto g_*(t)\exp(s\xi)$, is a curve from $g_*(t)$ to $g$ for some $\xi \in T_{g_*(t)}G$, and $V$ denotes the vector field along $c$, which is defined by using the parallel transport of $v \in T_{g_*(t)}G$ along $c$, i.e. $V(s) := \tau_{c(0)c(s)}v$ for $s \in [0,1]$.
This shows, that condition 2) can be replaced then by the assumption $\|D^2\tilde{\Omega}(g,t)\| \leq c_3$ for all $g \in \mathcal{B}_R(g_*(t))$.

**Time-varying Newton algorithm on Lie groups**
The update scheme, corresponding to the dynamical system (2.52), computes $g_*(t)$ at times $t_k = kh$ for $\mathcal{M} = -\frac{1}{h}$, $k \in \mathbb{N}$ and step size $h > 0$. It is given by

$$\boxed{g_{k+1} = \exp_{g_k}(hg_k\Omega(g_k,t_k)),} \tag{2.53}$$

where $\exp_{g_k}$ denotes the exponential map at $g_k$ and $\Omega(g_k,t_k)$ satisfies

$$\frac{1}{2}[\Omega(g_k,t_k),\tilde{\Omega}(g_k,t_k)]+D\tilde{\Omega}(g_k,t_k)\cdot g_k\Omega(g_k,t_k)=-\frac{1}{h}\tilde{\Omega}(g_k,t_k)-\tilde{\Omega}_\tau^h(g_k,t_k).$$

Here, $\tilde{\Omega}_\tau^h(g_k,t_k)$ denotes an approximation of $\frac{\partial}{\partial t}\tilde{\Omega}(g_k,t_k)$.
By applying Main Theorem 2.2 to the situation here, we immediately obtain the following result.

**Theorem 2.8.** *Let $G$ denote a Lie group endowed with a bi-invariant metric and let $R>0$ be any real number with $i^*(M)\geq R$ . Let $F:G\times\mathbb{R}\to TG$, $(g,t)\mapsto F(g,t)=: g\tilde{\Omega}(g,t)$ be a smooth vector field. Let $t\mapsto g_*(t)$ be a smooth isolated zero of $F$, i.e. $\tilde{\Omega}(g_*(t),t)=0$ for all $t$. Assume further the existence of constants $c_1,c_2,c_3,c_4,c_5,c_6,c_7>0$ such that*

*(i)* $\|DF(g_*(t),t)\|\leq c_1$, $\|\frac{\partial}{\partial t}F(g_*(t),t)\|\leq c_2$, $\|DF(g_*(t),t)^{-1}\|\leq c_3$ *for all* $t\in\mathbb{R}$,

*(ii)* $\|\nabla_{\dot{c}_{g_0}^g(s)}\nabla_{Y_w}F(c_{g_0}^g(s),t)\|_{c_{g_0}^g(s)}\leq c_4$ *for all* $g_0,g\in\mathcal{B}_R(g_*(t))$, $s\in[0,R_{g_0}^g]$ *and* $w\in T_gG$ *with* $\|w\|_g=1$, $t\in\mathbb{R}$.

*Here* $R_{g_0}^g:=\operatorname{dist}(g_0,g)$, $c_{g_0}^g:[0,R_{g_0}^g]\to G$ *is a geodesic from* $g_0$ *to* $g$ *and* $Y_w$ *is a smooth vector field on* $\mathcal{B}_R(g_*(t))$ *defined via the parallel transport by* $Y_w(g'):=\tau_{g_*(t)g'}\tau_{g_*(t)g}^{-1}w$.

*(iii)* $\|\frac{\partial^2}{\partial t^2}F(g,t)\|\leq c_5$, $\|\frac{\partial}{\partial t}DF(g,t)\|\leq c_6$ *for all* $g\in\mathcal{B}_R(g_*(t))$, $t\in\mathbb{R}$.

*(iv)* $\|\tilde{\Omega}_\tau^h(g,t)-\frac{\partial}{\partial t}\tilde{\Omega}(g,t)\|\leq c_7h$, *for all* $g\in\mathcal{B}_R(g_*(t))$, $t\in\mathbb{R}$, $h>0$.

*Then for $c>0$ and sufficiently small $h>0$, the sequence defined by* (2.53) *satisfies for* $k\in\mathbb{N}$

$$\operatorname{dist}(g_k,g_*(t_k))\leq ch$$

*provided that* $\operatorname{dist}(g_0,g_*(0))<ch$. *Thus, the update scheme* (2.53) *is well defined and produces estimates for $g_*(t_k)$, whose accuracy can be controlled by the step size.*

**Example 2.1. Square roots**
Let $G=(0,\infty)\subset\mathbb{R}$. Define a left-invariant Riemannian metric on $G$ by

$$\langle u,v\rangle_g:=\langle g^{-1}u,g^{-1}v\rangle_e:=mg^{-2}uv,\qquad m>0$$

and note, that it is also bi-invariant. Then let $a(t)>0$ for all $t$ and define $F(g,t)=g\tilde{\Omega}(g,t)=a(t)-g^2$. We get

$$\nabla_{\dot{g}}F(g,t)=\frac{1}{2}g[\Omega,\tilde{\Omega}]+gD\tilde{\Omega}(g,t)\dot{g}=$$

$$gD\tilde{\Omega}(g,t)\dot{g}=gD(g^{-1}a(t)-g)\dot{g}=g(-g^{-2}a(t)-1)\dot{g}=(-g-g^{-1}a(t))\dot{g}.$$

Thus the time-varying Newton flow on $G$ is given for $\mathcal{M}(g)=-\sigma$ ($\sigma>0$) by

$$(-g - g^{-1}a(t))\dot{g} = -\sigma(a(t) - g^2) - \dot{a}(t).$$

Or explicitly,

$$\dot{g} = \frac{g}{g^2 + a(t)}(\sigma a(t) - \sigma g^2 + \dot{a}(t)).$$

Note that the Newton flow in the ambient Euclidean space $\mathbb{R}$ is given by

$$DF(x,t) \cdot \dot{x} = -\sigma F(x,t) - \frac{\partial}{\partial t}F(x,t).$$

Thus we have

$$-2x\dot{x} = -\sigma a(t) + \sigma x^2 - \dot{a}(t),$$

which can be rewritten in explicit form by

$$\dot{x} = \frac{1}{2x}(\sigma a(t) - \sigma x^2 + \dot{a}(t)).$$

**Example 2.2. Cholesky factorization**
Let $A(t) \in \mathbb{R}^{n\times n}$ be a smooth family of symmetric positive definite matrices. Then for any $t \in \mathbb{R}$, there exists an unique lower triangular matrices $L \in \mathbb{R}^{n\times n}$ such that

$$A(t) = LL^\top.$$

Thus for $G = B_n := \left\{ \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ 0 & \dots & a_{nn} \end{bmatrix} \middle| \; a_{ii} > 0, i = 1, \dots, n \right\}$, the time-varying task is to determine $g(t) \in G$ such that for all $t$ holds

$$A(t) = g(t)^\top g(t).$$

We use the Riemannian metric

$$\langle \xi, \eta \rangle_g := \operatorname{tr}((g^{-1}\xi)^\top g^{-1}\eta) = \operatorname{tr}(\xi^\top (gg^\top)^{-1}\eta)$$

and note that it is left-invariant.
Let $\dot{g} = g\Omega$, $F(g,t) = g^\top g - A(t) =: g\tilde{\Omega}(g,t)$. To compute the affine connection, we want to use Theorem 2.5 and consider

$$\operatorname{tr}(((\operatorname{ad}x)^* y)^\top z) = \operatorname{tr}(y^\top [x,z]) = \operatorname{tr}(([y^\top, x]^\top)^\top z) = \operatorname{tr}([x^\top, y]^\top z).$$

Note that

$$[x^\top, y] = [x^\top, y]_{upper} + [x^\top, y]_{lower},$$

where $A_{upper}$ denotes the upper diagonal part including the diagonal of a square matrix $A$, while $A_{lower}$ denotes the lower diagonal part of $A$. Since $\operatorname{tr}(([x^\top, y]_{lower})^\top z) = 0$, we get that

$$\operatorname{tr}([x^\top, y]^\top z) = \operatorname{tr}(([x^\top, y]_{upper})^\top z),$$

which shows that

$$(\mathrm{ad}x)^* y = [x^\top, y]_{upper}.$$

Therefore, the connection form $\omega$ is given by $\omega(x,y) = \frac{1}{2}([x,y] - [x^\top, y]_{upper} - [y^\top, x]_{upper})$ and we obtain

$$\nabla_{\dot g} F(g,t) = \frac{1}{2} g([\Omega, \tilde\Omega] - [\Omega^\top, \tilde\Omega]_{upper} - [\tilde\Omega^\top, \Omega]_{upper}) + gD\tilde\Omega(g,t)\dot g =$$

$$\frac{1}{2} g([\Omega, g^{-1}(g^\top g - A(t))] - [\Omega^\top, g^{-1}(g^\top g - A(t))]_{upper} - [(g^\top g - A(t))^\top g^{-T}, \Omega]_{upper}) +$$

$$gD(g^{-1}(g^\top g - A(t)))\dot g.$$

Note that

$$D(g^{-1}(g^\top g - A(t)))\dot g = -g^{-1}\dot g g^{-1}(g^\top g - A(t)) + g^{-1}(\dot g^\top g + g^\top \dot g) =$$

$$-g^{-1} g\Omega g^{-1}(g^\top g - A(t)) + g^{-1}(\Omega^\top g^\top g + g^\top g\Omega).$$

Hence the time-varying Newton flow reads

$$\frac{1}{2} g([\Omega, g^{-1}(g^\top g - A(t))] - [\Omega^\top, g^{-1}(g^\top g - A(t))]_{upper} - [(g^\top g - A(t))^\top g^{-T}, \Omega]_{upper}) -$$

$$\Omega g^{-1}(g^\top g - A(t)) + g^{-1}(\Omega^\top g^\top g + g^\top g\Omega) = \mathcal{M}(g) g\tilde\Omega + \dot A(t).$$

**Example 2.3. Symmetric eigenvalue problem**
Let $A(t) \in \mathbb{R}^{n\times n}$ be a smooth family of symmetric matrices with distinct eigenvalues. To track the diagonalizing transformation $g_*(t_k) \in \mathrm{SO}(n)$ of $A(t_k)$ at discrete times $t_k = kh$ for $k \in \mathbb{N}$ and $h > 0$, we use the following approach.
Let $G = \mathrm{SO}(n)$, $N \in \mathbb{R}^{n\times n}$, $\langle g\xi, g\eta\rangle_g := \mathrm{tr}\,(\xi^\top \eta)$ for $g\xi, g\eta \in T_g\mathrm{SO}(n)$. Let

$$f : \mathrm{SO}(n) \times \mathbb{R} \to \mathbb{R},$$

defined by

$$f(g,t) := \mathrm{tr}\,(N g^\top A(t) g).$$

Then

$$Df(g,t)\cdot H = \mathrm{tr}\,(NH^\top A(t)g + Ng^\top A(t)H) = \mathrm{tr}\,(H^\top A(t)gN + H^\top A(t)gN) =$$

$$2\mathrm{tr}\,(H^\top A(t)gN),$$

and the intrinsic gradient with respect to $g$ is given by

$$\mathrm{grad}\, f(g,t) = \pi_{T_gG}(2A(t)gN) = 2g(g^\top A(t)gN)_{sk} = g(g^\top A(t)gN - Ng^\top A(t)g) =$$

$$g[g^\top A(t)g, N].$$

Note that for diagonal $N$ with distinct eigenvalues, the gradient $\mathrm{grad}\, f(g,t)$ is zero if and only if $g^\top A(t)g$ is diagonal. Thus the solution of the Newton flow tracks the diagonalizing orthogonal transformation of $A(t)$.

To compute the Newton flow, we use equation (2.49) and get for $F(g,t) = \mathrm{grad}\, f(g,t)$:

$$\nabla_{\dot g}\mathrm{grad}\, f(g,t) = g\left(\frac{1}{2}[g^\top \dot g, [g^\top A(t) g, N]] + D[g^\top A(t) g, N]\cdot \dot g\right).$$

By using the abbreviations $\Omega = g^\top \dot g$, $B = g^\top A(t) g$, this simplifies to

$$\nabla_{\dot g}\mathrm{grad}\, f(g,t) = \frac{1}{2} g[\Omega, [B,N]] + g[\dot g^\top A(t) g + g^\top A(t)\dot g, N] =$$

$$\frac{1}{2} g[\Omega, [B,N]] + g[[B,\Omega],N] = \frac{1}{2} g\big([[B,\Omega,]N] + [[N,\Omega],B]\big).$$

Therefore, the intrinsic Newton flow is given by

$$\frac{1}{2} g\big([[B,\Omega,]N] + [[N,\Omega],B]\big) = \mathcal{M}(g)\mathrm{grad}\, f(g,t) - \frac{\partial}{\partial t}\mathrm{grad}\, f(g,t). \tag{2.54}$$

**Approximative solution**

Let for $t \in \mathbb{R}$, $g_* \in \mathrm{SO}(n)$ such that $g_*^\top A(t) g_* =: D$ is diagonal. Thus, if $g \approx g_*$, then $B \approx D$ is almost a diagonal matrix. By approximating $B$ by its diagonal part $\tilde B$, one can explicitly solve (2.54). Hence, let

$$\tilde B_{ij} = \begin{cases} (g^\top A(t) g)_{ii}, & i = j \\ 0, & i \neq j. \end{cases}$$

and consider

$$[[\tilde B, \Omega,]N] + [[N,\Omega],\tilde B] = R, \tag{2.55}$$

to get a formula for $\Omega = -\Omega^\top$, where

$$R := 2g^\top\left(\mathcal{M}(g)\mathrm{grad}\, f(g,t) - \frac{\partial}{\partial t}\mathrm{grad}\, f(g,t)\right).$$

Equation (2.55) can be rewritten as

$$[\tilde B\Omega - \Omega\tilde B, N] + [N\Omega - \Omega N, \tilde B] = R,$$

or equivalently,

$$\tilde B\Omega N - \Omega\tilde B N - N\tilde B\Omega + N\Omega\tilde B + N\Omega\tilde B - \Omega N\tilde B - \tilde B N\Omega + \tilde B\Omega N = R.$$

Since $N$ and $\tilde B$ are diagonal, we can now give a formula for the entries of $\Omega$:

$$2(N_{jj} - N_{ii})(\tilde B_{ii} - \tilde B_{jj})\Omega_{ij} = R_{ij}. \tag{2.56}$$

Note that $R$ turns for $\mathcal{M}(g) = -\frac{1}{h}$ into

$$R := 2g^\top\left(-\frac{1}{h} g[g^\top A(t) g, N] - g[g^\top \dot A(t) g, N]\right) = -2[g^\top(\frac{1}{h}A(t) + \dot A(t))g, N].$$

Thus

$$R_{ij} = -2(g^\top(\frac{1}{h}A(t) + \dot{A}(t))g)_{ij}(N_{jj} - N_{ii})$$

and equation (2.56) uniquely determines the entries of the skew symmetric matrix $\Omega$

$$\Omega(g,t)_{ij} = \begin{cases} \frac{(g^\top(\frac{1}{h}A(t)+\dot{A}(t))g)_{ij}}{\tilde{B}_{jj}-\tilde{B}_{ii}}, & i \neq j \\ 0, & i = j. \end{cases}$$

By using this explicit expression of $\Omega$, the update scheme to track the desired orthogonal matrices $g_*(t_k)$ of the symmetric eigenvalue problem is given by

$$\boxed{g_{k+1} = \exp_{g_k}\left(hg_k\Omega(g_k,t_k).\right)}$$

This shows, that the extrinsic approach to solve the eigenvalue problem in Chapter 4.1 leads to an update scheme, which extends the formula here, since it differs by additional terms producing directions towards the manifold $\mathrm{SO}(n)$.

The previous example can be generalized to an arbitrary Lie group $G$ with Lie algebra $\mathfrak{g}$ and a left-invariant Riemannian metric $\langle,\rangle_g$ on $G$.
Choose a regular element $n$ in the Lie algebra $\mathfrak{g}$. Given a differentiable curve $a : I \to \mathfrak{g}$, we want to find $g(t) \in G$, $t \in I$, such that

$$[n, \mathrm{Ad}(g(t))a(t)] = 0 \quad \forall t \in I.$$

Thus we consider the differentiable map

$$F : G \times \mathbb{R} \to TG,$$

defined by

$$F(g,t) := L_g[n, \mathrm{Ad}(g(t))a(t)].$$

Then we get with the above notation and $\dot{g} = L_g\Omega$

$$\nabla_{\dot{g}}F + \frac{\partial}{\partial t}F =$$

$$L_g\left(\frac{1}{2}[\Omega,[n,\mathrm{Ad(g)}a(t)]] + [n,[Ad(g^{-1}a,\Omega]] + [n, Ad(g^{-1})\dot{a}]\right) =$$

$$\mathcal{M}(g)L_g([n,\mathrm{Ad(g^{-1})}a]),$$

which is equivalent to

$$\frac{1}{2}[\Omega,[n,\mathrm{Ad(g)}a(t)]] + [n,[\mathrm{Ad}(g^{-1}a,\Omega]] + [n,\mathrm{Ad}(g^{-1})\dot{a}] = L_g^{-1}\mathcal{M}(g)L_g([n,\mathrm{Ad(g^{-1})}a]).$$

Thus the task is to solve the linear equation for $\Omega$

$$\mathcal{F}(\Omega) = -[n,\mathrm{Ad}(g^{-1})\dot{a}] + L_g^{-1}\mathcal{M}(g)L_g([n,\mathrm{Ad(g^{-1})}a]),$$

where $\mathcal{F} : \mathfrak{g} \to \mathfrak{g}$ is the linear operator on the Lie algebra, defined by

$$\mathcal{F}(\Omega) = \frac{1}{2}[\Omega,[n,b]] + [n,[b,\Omega]].$$

Note that $\mathcal{F} : \mathfrak{g} \to \mathfrak{g}$ is invertible under suitable assumptions on $n$ and $a$ (such as defining regular elements in the Lie algebra for any $t$).

## 2.3 Euclidean time-varying Newton flow

In this chapter we consider the time-varying Newton flow in Euclidean space, which is the simplest special case of the previously introduced Riemannian algorithms. The resulting methods are therefore easier to understand, when applied to a particular problem. We also derive modifications of the tracking theorems, where inexact and underdetermined versions of the Newton flows are examined. It seems obvious to study these variants in order to increase the applicability and the performance of the proposed algorithms.

We consider a smooth map

$$F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n,$$

defined by

$$(x, t) \mapsto F(x, t).$$

Assume that there exists a smooth zero of $F$, i.e. a curve $x_* : \mathbb{R} \to \mathbb{R}^n$ such that $F(x_*(t), t) = 0$ holds for all $t \in \mathbb{R}$. In order to determine this zero for all $t$, we use the time varying Newton flow (2.12), which is given in this situation by

$$DF(x,t) \cdot \dot{x} + \frac{\partial}{\partial t} F(x,t) = \mathcal{M}(x) F(x,t). \tag{2.57}$$

Here, $DF$ denotes the "usual" derivative in $\mathbb{R}^n$ and $\mathcal{M}$ is a stable bundle map, cf. Definition 2.3.
To rewrite the above differential equation in an explicit form, we assume the existence of $r > 0$ such that $\mathrm{rk} DF(x, t) = n$, for all $t \in \mathbb{R}$ and $x \in B_r(x_*(t))$. Then (2.57) is equivalent to

$$\boxed{\dot{x} = DF(x,t)^{-1} \left( \mathcal{M}(x) F(x,t) - \tfrac{\partial}{\partial t} F(x,t) \right).} \tag{2.58}$$

This assumption moreover implies that the smooth zero $x_*(t)$ of $F(x, t)$ is isolated, i.e. for $x \in B_r(x_*(t))$ holds: $F(x, t) = 0$ if and only if $x = x_*(t)$.

### 2.3.1 The tracking algorithms in Euclidean space

At first we consider the continuous case and formulate the Euclidean version of the Riemannian tracking Main Theorem 2.1. Note that the assumptions for the higher order derivatives now turn into a quite simple form, compared to the original conditions.

**Theorem 2.9.** *Let $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ be a smooth map and $t \mapsto x_*(t)$ be a smooth isolated zero of $F$, i.e. $F(x_*(t), t) = 0$ for all $t \in \mathbb{R}$. Assume further there exist constants $R, L_1, L_2, L_3 > 0$ such that for all $t \in \mathbb{R}$ holds:*

1. $\|DF(x_*(t), t)\| \leq L_1$,
2. $\|DF(x_*(t), t)^{-1}\| \leq L_2$,

*3.* $\|D^2F(x,t)\| \leq L_3,$ *for all* $\|x - x_*(t)\| \leq R$, $t \geq 0$.

*Then there exists* $0 < r < R$ *such that for any initial condition* $x(0)$ *with* $\|x(0) - x_*(0)\| \leq r$ *there exists a unique solution* $x(t)$, $t \geq 0$ *of* (2.58) *with the properties*

1. $\|x(t) - x_*(t)\| \leq R$ *for all* $t \geq 0$.

2. $\|x(t) - x_*(t)\|$ *converges exponentially to* $0$.

**Euclidean time-varying Newton algorithm**
The discretization scheme introduced in the previous section for the tracking algorithm on Riemannian manifolds can be analogously employed in Euclidean space.
Thus the discrete version of (2.58) at discrete times $t_k = kh$ for $\mathcal{M}(x) = -\frac{1}{h}I$, $k \in \mathbb{N}$ and $h > 0$ is given by

$$\boxed{x_{k+1} = x_k - DF(x_k,t_k)^{-1}\left(F(x_k,t_k) + hF_\tau^h(x_k,t_k)\right).} \tag{2.59}$$

As done in the previous chapters, we used an approximation $F_\tau^h(x,t)$ of $\frac{\partial}{\partial t}F(x,t)$, cf. Section 2.1.2.2. The following result follows directly from Main Theorem 2.2 and gives conditions guaranteeing that the above sequence tracks the smooth zero $x_*(t)$ of $F(x,t)$ at discrete times $t_k$.

**Theorem 2.10.** *Let* $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$, $(x,t) \mapsto F(x,t)$ *be a smooth map and let* $t \mapsto x_*(t)$ *be a smooth isolated zero of* $F$*, i.e.* $F(x_*(t),t) = 0$ *for all* $t \in \mathbb{R}$*. Let further there exist constants* $c_1, c_2, c_3, c_4, c_5, c_6, c_7, R > 0$ *such that*

*(i)* $\|DF(x_*(t),t)\| \leq c_1$, $\|\frac{\partial}{\partial t}F(x_*(t),t)\| \leq c_2$, $\|DF(x_*(t),t)^{-1}\| \leq c_3$ *for all* $t \in \mathbb{R}$,

*(ii)* $\|D^2F(x,t)\| \leq c_4$, $\|\frac{\partial^2}{\partial t^2}F(x,t)\| \leq c_5$, $\|\frac{\partial}{\partial t}DF(x,t)\| \leq c_6$ *for all* $x \in B_R(x_*(t))$, $t \in \mathbb{R}$.

*(iii)* $\|F_\tau^h(x,t) - \frac{\partial F}{\partial t}(x,t)\| \leq c_7h$, *for all* $x \in B_R(x_*(t))$, $t \in \mathbb{R}$, $h > 0$.

*Then the following statements hold*

1. *There exist* $0 < r < R$ *and* $c_8, c_9, c_{10} > 0$ *such that for* $t \in \mathbb{R}$

$$\|x - x_*(t)\| \leq c_8\|F(x,t)\|, \tag{2.60}$$

$$\|F(x,t)\| \leq c_9\|x - x_*(t)\|, \tag{2.61}$$

*and*

$$\|DF(x,t)^{-1}\| \leq c_{10} \tag{2.62}$$

*for* $x \in B_r(x_*(t))$.

2. *The discretization sequence defined in* (2.59) *with* $t_k = kh$, $h > 0$ *satisfies for some* $c_{11}, c_{12} > 0$

$$\|x_{k+1} - x_*(t_{k+1})\| \le c_{11}\|x_k - x_*(t_k)\|^2 + c_{12}h^2 \tag{2.63}$$

*for* $x_k \in B_r(x_*(t_k))$, $k \in \mathbb{N}_0$

3. *Let* $c > 0$ *be constant and* $h$ *sufficiently small. For any initial condition* $x_0$ *with* $\|x_0 - x_*(0)\| < ch$ *we have*

$$\|x_k - x_*(t_k)\| \le ch$$

*for all* $k \in \mathbb{N}_0$. *Thus, the update scheme* (2.59) *is well defined and produces estimates for* $x_*(t_k)$, *whose accuracy can be controlled by the step size.*

It follows, that one can control the global tracking error of the discretization scheme by varying the step size $h$. Note that, in general, one cannot expect more than such a bound of the tracking error being hold, if one discretizes the Newton flow with a fixed step size. In particular, one cannot in general guarantee that the tracking error vanishes asymptotically for $k \to \infty$, as the next example shows.

**Example 2.4.** Let $F(x,t) := x - \sin t$. Then $x_*(t) := \sin t$ is a smooth isolated zero of $F(x,t)$, $DF(x,t) = \mathrm{id}_x$ and $\frac{\partial F}{\partial t}(x,t) = -\cos t$. Hence, a Newton flow is given by

$$\dot{x} = -x + \sin t + \cos t.$$

Let $t_k = 2\pi l$, for some $k, l \in \mathbb{N}$, $h > 0$ and $x_k = x_*(t_k) = \sin t_k = 0$. Then

$$x_{k+1} = x_k - (x_k + \sin t_k + \cos t_k)h = -\cos(2\pi l)h = -h,$$

and therefore

$$\|x_*(t_k + h) - x_{k+1}\| = \|x_*(2\pi l + h) - h\| =$$
$$\|\sin(2\pi l + h) - h\| = \|\sin(h) - h\| \ge \frac{h^3}{6}.$$

### 2.3.2 Inexact time-varying Newton flow

The tracking of a smooth isolated zero of a map $F$ by using the previously proposed algorithms, requires either to invert a matrix $DF$ or alternatively, to solve a linear system

$$DF \cdot \dot{x} = r$$

for $\dot{x} \in \mathbb{R}^n$ at each step for some $r \in \mathbb{R}^n$. As the size of $DF$ increases quadratically with the dimension $n$ of the image of $F$, these procedures are inappropriate for large $n$. But the computational effort can be reduced significantly, if one only solves the associated linear system approximatively. In Chapter 4.1 and Chapter 4.3, there are two situations, where useful approximations for the inverse of $DF$ are available. This

motivated to provide a theoretical basis for such approaches. Note that this idea is closely related to the so-called inexact Newton method, where zeros of time-invariant maps are determined by using a discrete Newton-type algorithm, cf. [19].
Hence, we now study the convergence properties of the approximative Newton flow of a smooth map

$$F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n,$$

defined by

$$(x, t) \mapsto F(x, t).$$

Again, we assume that $F$ has a smooth isolated zero $x_*(t)$, i.e. $t \mapsto x_*(t)$ is a smooth curve and $F(x_*(t), t) = 0$ for all $t \in \mathbb{R}$.
At first, we consider the perturbed time-varying Newton flow

$$DF(x, t)\dot{x} + \frac{\partial}{\partial t}F(x, t) = \mathcal{M}(x)F(x, t) + \Pi(x, t), \tag{2.64}$$

where $\mathcal{M}$ is a stable bundle map and $\Pi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ denotes the perturbation. The next lemma gives conditions such that the norm of $F(x(t), t)$ converges exponentially to zero, where $x(t)$ is a solution of the above ODE.

**Lemma 2.6.** *Let $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ and $\Pi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ as above and assume the existence of $r, c > 0$ such that*

$$\|\Pi(x, t)\| \leq c\|F(x, t)\|,$$

*for all $x \in B_r(x_*(t))$, $t \in \mathbb{R}$. Then for some $a, b > 0$, any solution $x(t)$ of (2.64) satisfies*

$$\|F(x(t), t)\| \leq ae^{-bt}, \tag{2.65}$$

*provided that $\mathcal{M}$ is a stable bundle map s.th. $\lambda := \sup_{v \in \mathbb{R}^n, \|v\|=1} \langle \mathcal{M}(x) \cdot v, v\rangle$ satisfies $\lambda < -c$.*

*Proof.*

$$\frac{d}{dt}\|F(x, t)\|^2 = 2\langle \frac{d}{dt}F(x, t), F(x, t)\rangle = 2\langle DF(x, t) \cdot \dot{x} + \frac{\partial}{\partial t}F(x, t), F(x, t)\rangle =$$

$$2\langle \mathcal{M}(x)F(x, t) + \Pi(x, t), F(x, t)\rangle = 2(\langle \mathcal{M}(x)F(x, t), F(x, t)\rangle + \langle \Pi(x, t), F(x, t)\rangle).$$

Thus

$$\frac{d}{dt}\|F(x, t)\|^2 \leq 2\|F(x, t)\|^2(\lambda + c), \tag{2.66}$$

where $\lambda = \sup_{v \in \mathbb{R}^n, \|v\|=1} \langle \mathcal{M}(x) \cdot v, v\rangle < 0$. Thus,

$$\|F(x, t)\| \leq \|F(x(0), 0)\|e^{(c+\lambda)t}.$$

□

We want to profit from this fact to deduce computationally easier systems to track $x_*(t)$.
If $F$ satisfies the conditions of Theorem 2.9, $DF(x,t)$ is invertible for $t \in \mathbb{R}$ and $x \in B_R(x_*(t))$ for some $R > 0$ and equation (2.64) can be rewritten as

$$\dot{x} = DF(x,t)^{-1}\left(\mathcal{M}(x)F(x,t) - \frac{\partial}{\partial t}F(x,t)\right) + DF(x,t)^{-1}\Pi(x,t). \tag{2.67}$$

Hence, if we have an approximation $G(x,t) : \mathbb{R}^n \to \mathbb{R}^n$ of $DF(x,t)^{-1}$ satisfying

$$DF(x,t)G(x,t)\left(\mathcal{M}(x)F(x,t) - \frac{\partial}{\partial t}F(x,t)\right) = \mathcal{M}(x)F(x,t) - \frac{\partial}{\partial t}F(x,t) + \Pi(x,t),$$

then

$$\boxed{\dot{x} = G(x,t)\left(\mathcal{M}(x)F(x,t) - \tfrac{\partial}{\partial t}F(x,t)\right)} \tag{2.68}$$

is equivalent to (2.67) and we can extend Theorem 2.9 for inexact time-varying Newton flows.

**Theorem 2.11.** *Let $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ be a smooth map and $t \mapsto x_*(t)$ be a continuously differentiable isolated zero of $F$, i.e. $F(x_*(t),t) = 0$ for all $t \in \mathbb{R}$. Assume further that there exist constants $R, L_1, L_2, L_3, L_4 > 0$ and an approximation $G(x,t) : \mathbb{R}^n \to \mathbb{R}^n$ for $DF(x,t)^{-1}$ such that for all $t \geq 0$ holds*

1. $\|DF(x_*(t),t)\| \leq L_1$,
2. $\|DF(x_*(t),t)^{-1}\| \leq L_2$,
3. $\|D^2F(x,t)\| \leq L_3$, *for all* $\|x - x_*(t)\| \leq R$,
4. $\left\|(DF(x,t)G(x,t) - I)\left(\mathcal{M}(x)F(x,t) - \frac{\partial}{\partial t}F(x,t)\right)\right\| \leq L_4\|F(x,t)\|$, *for all* $\|x - x_*(t)\| \leq R$.

*Then there exists $0 < r < R$ such that for any initial condition $x(0)$ with $\|x(0) - x_*(0)\| \leq r$ there exists a unique solution $x(t)$, $t \geq 0$ of (2.68) with the properties*

1. $\|x(t) - x_*(t)\| \leq R$ *for all* $t \geq 0$,
2. $\|x(t) - x_*(t)\|$ *converges exponentially to* $0$,

*provided that* $\lambda := \sup_{v\in\mathbb{R}^n,\|v\|=1} \langle \mathcal{M}(x) \cdot v, v\rangle < 0$ *satisfies* $\lambda < -L_4$.

*Proof.* Due to Proposition (2.5.6) of [1], p. 119, there exists a $0 < \hat{r} \leq R$ such that for any $t$ the map $x \mapsto F(x,t)$ is a diffeomorphism on $U_{\hat{r}}(x_*(t)) := \{x | \; \|x - x_*(t)\| \leq \hat{r}\}$. This implies the existence of constants $L_1',\ L_2' > 0$ such that

$$\|DF(x,t)\| \leq L_1',$$

$$\|DF(x,t)^{-1}\| \leq L_2',$$

for all $\|x - x_*(t)\| \leq \hat{r}$, $t \geq 0$.
From the mean value theorem we conclude the existence of some $k_1, k_2 > 0$ such that for $t \geq 0$

$$k_1\|x - x_*(t)\| \leq \|F(x,t) - F(x_*(t),t)\| \leq k_2\|x - x_*(t)\| \tag{2.69}$$

holds for all $x \in U_{\hat{r}}(x_*(t))$. Let $x(t)$ denote the maximal solution of (2.68) for $0 \leq t < t_+$ and $T := \sup\{t < t_+ | \; \|x(t) - x_*(t)\| \leq \hat{r}\}$. Let $r := \min\{\frac{k_1\hat{r}}{k_2}, \hat{r}\}$ and assume $T < \infty$. Note that $t \mapsto \|F(x(t),t)\|$ is strictly monotonically decreasing, since

$$\begin{aligned} \frac{d}{dt}\|F(x,t)\|^2 &= 2\left\langle DF(x,t)\dot{x} + \frac{\partial}{\partial t}F(x,t), F(x,t)\right\rangle \\ &= 2\left\langle DF(x,t)G(x,t)\left(\mathcal{M}F(x,t) - \frac{\partial}{\partial t}F(x,t)\right) + \frac{\partial}{\partial t}F(x,t), F(x,t)\right\rangle \\ &= 2\left\langle (DF(x,t)G(x,t) - I)\left(\mathcal{M}F(x,t) - \frac{\partial}{\partial t}F(x,t)\right) + \mathcal{M}F(x,t), F(x,t)\right\rangle \\ &\leq 2L_4\|F(x,t)\|^2 + 2\lambda\|F(x,t)\|^2 = 2(L_4+\lambda)\|F(x,t)\|^2 < 0, \end{aligned}$$

where we used assumption 4 and (2.68).
Therefore

$$\|x(t) - x_*(t)\| \leq \frac{1}{k_1}\|F(x(t),t)\| < \frac{1}{k_1}\|F(x(0),0)\| \leq \frac{k_2}{k_1}\|x(0) - x_*(0)\| \leq \hat{r}, \tag{2.70}$$

for all $0 \leq t < T$, $\|x(0) - x_*(0)\| \leq r$. Thus $\|x(T) - x_*(T)\| < \hat{r}$, contradicting the assumed finiteness of $T$. This shows that (2.70) holds for all $t \geq 0$, provided $\|x(0) - x_*(0)\| \leq r$. In particular, the solution $x(t)$ exists for all $t \geq 0$. Since $\|F(x(t),t)\|$ converges exponentially to 0, cf. Lemma 2.6, this implies the exponential convergence of $x(t)$ to $x_*(t)$.

□

**Inexact time-varying Newton algorithm**
We now consider the Euler discretization of (2.68), which yields a sequence approximating the exact zero of $F$ at discrete times $t_k = kh$ for $h > 0$ and $k \in \mathbb{N}$. The discrete tracking algorithm is given by

$$\boxed{x_{k+1} = x_k - G(x_k,t_k)\left(F(x_k,t_k) + hF_\tau^h(x_k,t_k)\right),} \tag{2.71}$$

where we formulated the update scheme by using an approximation $F_\tau^h(x_k,t_k)$ of $\frac{\partial}{\partial t}F(x_k,t_k)$. Note that we again set $\mathcal{M}(x) = -\frac{1}{h}I$, which is crucial to prove the stability result for (2.71) in the next theorem.

**Theorem 2.12.** *Let $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$, $(x,t) \mapsto F(x,t)$ be smooth. Let $t \mapsto x_*(t)$ be a smooth isolated zero of $F$, i.e. $F(x_*(t),t) = 0$ for all $t \in \mathbb{R}$. Let further there exist constants $c_1, c_2, c_3, c_4, c_5, c_6, c_7, R > 0$ such that*

(i) $\|DF(x_*(t),t)\| \le c_1$, $\|\frac{\partial}{\partial t}F(x_*(t),t)\| \le c_2$, $\|DF(x_*(t),t)^{-1}\| \le c_3$ *for all* $t \in \mathbb{R}$,

(ii) $\|D^2F(x,t)\| \le c_4$, $\|\frac{\partial^2}{\partial t^2}F(x,t)\| \le c_5$, $\|\frac{\partial}{\partial t}DF(x,t)\| \le c_6$ *for all* $x \in B_R(x_*(t))$, $t \in \mathbb{R}$.

(iii) $\|F_\tau^h(x,t) - \frac{\partial F}{\partial t}(x,t)\| \le c_7 h$, *for all* $x \in B_R(x_*(t))$, $t \in \mathbb{R}$, $h > 0$.

*Let further* $G(x,t)$ *denote an approximation for* $DF(x,t)^{-1}$ *satisfying for some* $c, \tilde{c} > 0$

$$\left\|(DF(x,t)G(x,t) - I)\left(\frac{1}{h}F(x,t) + F_\tau^h(x,t)\right)\right\| \le \tilde{c}\|F(x,t)\|, \qquad (2.72)$$

*for all* $h > 0$ *and* $x \in B_{ch}(x_*(t))$, $t \in \mathbb{R}$.
*Then the following statements hold*

1. *There exists* $0 < r \le R$ *and* $c_8, c_9, c_{10} > 0$ *such that for* $t \in \mathbb{R}$

$$\|x - x_*(t)\| \le c_8\|F(x,t)\|, \qquad (2.73)$$

$$\|F(x,t)\| \le c_9\|x - x_*(t)\|, \qquad (2.74)$$

   *and*

$$\|DF(x,t)^{-1}\| \le c_{10} \qquad (2.75)$$

   *for* $x \in B_r(x_*(t))$.

2. *The discretization sequence* $(x_k)$ *as defined in* (2.71) *with* $t_k = kh$, $h > 0$ *satisfies for some* $c_{11}, c_{12} > 0$

$$\|x_{k+1} - x_*(t_{k+1})\| \le c_{11}\|x_k - x_*(t_k)\|^2 + c_{12}h^2 \qquad (2.76)$$

   *for* $x_k \in B_r(x_*(t_k))$ *with* $r \le ch$, $k \in \mathbb{N}_0$.

3. *Let* $c > 0$ *be constant and* $h$ *sufficiently small. For any initial condition* $x_0$ *with* $\|x_0 - x_*(0)\| < ch$ *we have*

$$\|x_k - x_*(t_k)\| \le ch$$

   *for all* $k \in \mathbb{N}_0$. *Thus, the update scheme* (2.71) *is well defined and produces estimates for* $x_*(t_k)$, *whose accuracy can be controlled by the step size.*

*Proof.* This theorem is very similar to Theorem 2.10 such that claim 1) and 3) follow from there. However, the proof of 2) differs by additionally perturbation terms.

We show (2.76) by bounding the norm of $F$ at $t = t_{k+1}$. Taylor's Theorem shows that

$$F(x_{k+1}, t_{k+1}) = F(x_k, t_k) + \frac{\partial F}{\partial t}(x_k, t_k)h + DF(x_k, t_k)(x_{k+1} - x_k) + \mathcal{R}, \qquad (2.77)$$

where $\mathcal{R}$ satisfies for $\Delta := \|x_{k+1} - x_k\|$

$$\|\mathcal{R}\| \le (c_4\Delta^2 + c_5h^2 + c_6\Delta h) \le \left(c_4 + \frac{c_6}{2}\right)\Delta^2 + \left(c_5 + \frac{c_6}{2}\right)h^2.$$

Using the update scheme (2.71) to replace $x_{k+1}$, (2.77) turns into

$$F(x_{k+1}, t_{k+1}) = \tag{2.78}$$
$$F(x_k, t_k) + \frac{\partial F}{\partial t}(x_k, t_k)h + DF(x_k, t_k)\left(hG(x_k, t_k)\left(-\frac{1}{h}F(x_k, t_k) - F_\tau^h(x_k, t_k)\right)\right) + \mathcal{R}.$$

Estimate (2.72) implies that

$$DF(x,t)G(x,t)\left(\frac{1}{h}F(x,t) + F_\tau^h(x,t)\right) = \left(\frac{1}{h}F(x,t) + F_\tau^h(x,t)\right) - \Pi(x,t),$$

for some perturbation term $\Pi(x,t)$ satisfying $\|\Pi(x,t)\| \leq \tilde{c}\|F(x,t)\|$. Thus (2.78) is equivalent to

$$F(x_{k+1}, t_{k+1}) = \frac{\partial F}{\partial t}(x_k, t_k)h - hF_\tau^h(x_k, t_k) + h\Pi(x_k, t_k) + \mathcal{R},$$

and therefore

$$\|F(x_{k+1}, t_{k+1})\| \leq c_7 h^2 + \left(c_4 + \frac{c_6}{2}\right)\Delta^2 + \left(c_5 + \frac{c_6}{2}\right)h^2 + h\tilde{c}\|F(x_k, t_k)\|$$

implying that

$$\|F(x_{k+1}, t_{k+1})\| \leq \left(c_4 + \frac{c_6}{2}\right)\Delta^2 + \left(c_5 + \frac{c_6}{2} + c_7\right)h^2 + \tilde{c}c_9 h\|x_k - x_*(t_k)\|$$

and

$$\|F(x_{k+1}, t_{k+1})\| \leq \left(c_4 + \frac{c_6}{2}\right)\Delta^2 + \left(c_5 + \frac{c_6 + \tilde{c}c_9}{2} + c_7\right)h^2 + \frac{\tilde{c}c_9}{2}\|x_k - x_*(t_k)\|^2.$$

Using (2.73) shows that

$$\|x_{k+1} - x_*(t_{k+1})\| \leq \tag{2.79}$$
$$c_8\left(c_4 + \frac{c_6}{2}\right)\Delta^2 + c_8\left(c_5 + \frac{c_6 + \tilde{c}c_9}{2} + c_7\right)h^2 + \frac{\tilde{c}c_8c_9}{2}\|x_k - x_*(t_k)\|^2.$$

We now inspect $\Delta$ and get

$$\Delta = \|x_{k+1} - x_k\| = \left\|hG(x_k, t_k)\left(\frac{1}{h}F(x_k, t_k) + F_\tau^h(x_k, t_k)\right)\right\| \leq$$
$$\left\|DF(x_k, t_k)^{-1}\big(F(x_k, t_k) + hF_\tau^h(x_k, t_k) + h\Pi(x_k, t_k)\big)\right\|.$$

Therefore,

$$\Delta \leq c_{10}(\|F(x_k, t_k)\| + h\|F_\tau^h(x_k, t_k)\| + \tilde{c}h\|F(x_k, t_k)\|). \tag{2.80}$$

Note that $\|F_\tau^h(x_k, t_k)\| \leq \|\frac{\partial}{\partial t}F(x_k, t_k)\| + c_7 h \leq c_2 + c_6\|x_k - x_*(t_k)\| + c_7 h$. Thus (2.80) turns into

$$\Delta \leq c_{10}\big(\|F(x_k, t_k)\| + h(c_2 + c_6\|x_k - x_*(t_k)\| + c_7 h) + \tilde{c}h\|F(x_k, t_k)\|\big).$$

It follows that

$$\Delta \leq c_{10}\big(c_9\|x_k - x_*(t_k)\| + h(c_2 + c_6\|x_k - x_*(t_k)\| + c_7 h) + \tilde{c}c_9 h\|x_k - x_*(t_k)\|\big) \leq$$

$$c_9 c_{10}\|x_k - x_*(t_k)\| + c_{10}(c_2 + (c_6 + \tilde{c}c_9)r + c_7 h)h.$$

By using the abbreviations $k_1 = c_9 c_{10}$ and $k_2 = c_{10}(c_2 + (c_6 + \tilde{c}c_9)r + c_7 h)$, this equation implies that

$$\Delta^2 \leq k_1^2\|x_k - x_*(t_k)\|^2 + k_2^2 h^2 + 2k_1 k_2\|x_k - x_*(t_k)\|h,$$

and hence

$$\Delta^2 \leq (k_1^2 + k_1 k_2)\|x_k - x_*(t_k)\|^2 + (k_2^2 + k_1 k_2)h^2.$$

Plug this into (2.79) and obtain

$$\|x_{k+1} - x_*(t_{k+1})\| \leq \big(k_3(k_1^2 + k_1 k_2) + k_4\big)\,\|x_k - x_*(t_k)\|^2 + \big(k_3 k_2^2 + k_1 k_2 k_3 + k_5\big)\,h^2,$$

where $k_3 = c_8\,(c_4 + c_6/2)$, $k_4 = \frac{\tilde{c}c_8 c_9}{2}$ and $k_5 = c_8\left(c_5 + c_7 + \frac{c_6 + \tilde{c}c_9}{2}\right)$.

□

### 2.3.3 Underdetermined Newton flow

We now derive a tracking algorithm for time-varying zeros of non-invertible maps in Euclidean space. This merges the Newton flow for underdetermined constant linear systems, as studied by Tanabe [66], with the Euclidean time-varying Newton flow introduced in this work. At the end of this paragraph, a discrete algorithm will be given, which also includes the inexact case, i.e. we define an update scheme, which is a discretization of an approximative underdetermined Newton flow. Analogously to the previous chapters, the stability of the resulting tracking algorithm will be shown.
We consider a smooth map

$$F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^m,$$

defined for $m < n$ by

$$(x, t) \mapsto F(x, t).$$

For $t \in \mathbb{R}$, let $\mathcal{X}(t) := \{x \in \mathbb{R}^n |\ F(x, t) = 0\}$ denote the zero set of $F$ and assume that it is not empty. Assume further that

$$\mathrm{rk} DF(x, t) = m$$

for all $x \in \mathcal{X}(t)$, $t \in \mathbb{R}$. Then 0 is a regular value of $x \mapsto F(x, t)$ and $\mathcal{X}(t) \subset \mathbb{R}^n$ is a Riemannian submanifold of dimension $n - m$, $t \in \mathbb{R}$.
Moreover, $\mathrm{rk}\big(DF(x, t)\ \frac{\partial}{\partial t}F(x, t)\big) = m$ for all $x \in \mathcal{X}(t)$, $t \in \mathbb{R}$, showing that $\hat{N} := F^{-1}(0)$ is a smooth Riemannian submanifold of $\mathbb{R}^{n+1}$ of dimension $n - m + 1$. Obviously, $\hat{N} = \{(x, t)|\ x \in \mathcal{X}(t),\ t \in \mathbb{R}\}$. Thus the zero sets considered here are smoothly changing time-varying manifolds instead of isolated curves. This expression is clarified in the next definition.

**Definition 2.5.** *Let $M(t) \subset \mathbb{R}^n$ be a family of $k$-dimensional Riemannian submanifolds for $t \in I$. We call $M(t)$ a smooth time-varying manifold, if the set*

$$\hat{M} := \{(x,t) |\ x \in M(t), t \in I\}$$

*is a smooth Riemannian submanifold of $\mathbb{R}^{n+1}$ for all open intervals $I \subset \mathbb{R}$.*

It is obvious, that the family of manifolds $\mathcal{X}(t)$ as defined above, is a smooth time-varying manifold. The goal is to construct a dynamical system, whose solution $x(t)$ converges exponentially to a connected component $\mathcal{X}_*(t)$ of $\mathcal{X}(t)$, i.e.

$$\operatorname{dist}(x(t), \mathcal{X}_*(t)) \leq ae^{-bt},$$

for some $a, b > 0$. Thus, we consider again the time-varying Newton flow

$$DF(x,t) \cdot \dot{x} + \frac{\partial}{\partial t} F(x,t) = \mathcal{M}(x) F(x,t), \tag{2.81}$$

where $\mathcal{M}$ is a stable bundle map. If there exists a $R > 0$ such that $\operatorname{rk} DF(x,t) = m$ for all $x \in U_R(t) := \{x \in \mathbb{R}^n |\ \operatorname{dist}(x, \mathcal{X}_*(t)) < R\}$, $t \in \mathbb{R}$, we can locally determine $DF(x,t)^\dagger : \mathbb{R}^m \to \mathbb{R}^n$, such that

$$DF(x,t) DF(x,t)^\dagger = I_m,$$

and we then call $DF(x,t)^\dagger$ a pseudo-inverse of $DF(x,t)$. Note that such operators exist under the above conditions, e.g. the Moore-Penrose inverse is given as

$$DF(x,t)^\dagger := DF(x,t)^\top (DF(x,t) DF(x,t)^\top)^{-1} \in \mathbb{R}^{n \times m}.$$

Moreover, the largest singular value of a pseudo-inverse $DF(x,t)^\dagger$ satisfies for $t \in \mathbb{R}$ and $x \in \mathbb{R}^n$

$$\sigma_{\max}(DF(x,t)^\dagger) = 1/\sigma_{\min}(DF(x,t)).$$

Using a pseudo inverse, a solution of (2.81) can be found by determining a solution $x(t)$ to

$$\dot{x} = DF(x,t)^\dagger \left( \mathcal{M}(x) F(x,t) - \frac{\partial}{\partial t} F(x,t) \right), \tag{2.82}$$

as any solution of (2.82) satisfies (2.81). Moreover, any solution $x(t)$ of (2.81) (and hence any solution of (2.82)) satisfies for all $t \in \mathbb{R}$

$$\|F(x(t),t)\| \leq ae^{-bt}, \tag{2.83}$$

for some $a, b > 0$ since $\frac{d}{dt} F(x(t),t) = \mathcal{M}(x(t)) F(x(t),t)$.
To derive the tracking algorithm, we have to be able to estimate the distance of a point to the zero set of $F$ by using the norm of the map $F$. This can be achieved by using the next lemma.

**Lemma 2.7.** *Let $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^m$ and $\mathcal{X}_*(t)$ as above. Let there exist $M, R, S, s > 0$ such that the following statements hold:*

1. *For all* $x \in \mathcal{X}_*(t)$, $t \in \mathbb{R}$, *the singular values* $\sigma_1(x,t), ..., \sigma_m(x,t)$ *of* $DF(x,t)^\top$ *satisfy*
$$s \leq \sigma_1(x,t), ..., \sigma_m(x,t) \leq S.$$
2. $\|D^2F(x,t)\| \leq M$ *for all* $x \in U_R(t) := \{x \in \mathbb{R}^n | \operatorname{dist}(x, \mathcal{X}_*(t)) < R\}$, $t \in \mathbb{R}$.

*Then there exist constants* $\kappa_1, \kappa_2 > 0$ *such that for* $r = \min\{R, \frac{s}{2M}\}$ *and all* $x \in U_r(t) = \{x \in \mathbb{R}^n | \operatorname{dist}(x, \mathcal{X}_*(t)) < R\}$, $t \in \mathbb{R}$ *holds*

*(i)* $\kappa_1 \operatorname{dist}(x, \mathcal{X}_*(t)) \leq \|F(x,t)\| \leq \kappa_2 \operatorname{dist}(x, \mathcal{X}_*(t))$,

*(ii)* $\sigma_{\min}(DF(x,t)^\top) \geq \frac{s}{2}$.

*Proof.* The right side of claim $(i)$ is obvious due to assumption 1) and 2). Therefore let for $t \in \mathbb{R}$, $x_* \in \mathcal{X}_*(t)$ denote a point of minimal distance to $x$, where $x \in U_r(t)$. Thus,

$$\operatorname{dist}(x, \mathcal{X}_*(t)) = \|x - x_*\|$$

and the vector $h := x - x_*$ is in the normal space of $\mathcal{X}_*(t)$ at $x_*$ and is orthogonal to the kernel of $DF(x_*, t)$. Thus,

$$\|DF(x_*, t) \cdot h\| \geq s\|h\|.$$

From Taylor's Theorem, we have

$$F(x,t) = DF(x_*, t) \cdot h + R_1,$$

where $R_1 = \int_0^1 D^2F(x_* + \tau h, t) \cdot (h,h) d\tau$. Therefore

$$\|F(x,t)\| \geq \|DF(x_*, t) \cdot h\| - \|R_1\|$$
$$\geq s\|h\| - M\|h\|^2 = \|h\|(s - M\|h\|).$$

Thus $(i)$ holds with $\kappa_1 = s/2$, since $\|h\| = \operatorname{dist}(x, \mathcal{X}_*(t)) \leq r \leq \frac{s}{2M}$.
To prove $(ii)$, let $v \perp \ker DF(x,t)$ and consider

$$DF(x,t)v = DF(x_*, t)v + \int_0^1 D^2F(x_* + \tau w, t) \cdot (w, v) d\tau,$$

where $w = x - x_*$. Thus

$$\|DF(x,t)v\| = \left\| DF(x_*, t)v + \int_0^1 D^2F(x_* + \tau w, t) \cdot (w, v) d\tau \right\| \geq$$
$$\geq \|DF(x_*, t)v\| - \left\| \int_0^1 D^2F(x_* + \tau w, t) \cdot (w, v) d\tau v \right\| \geq (s - M\|w\|)\|v\|.$$

Therefore, $\sigma_{\min}(DF(x,t)^\top) \geq s/2$ for $\|w\| \leq r$. □

We now have the necessary tools to prove the following result.

**Theorem 2.13.** *Let $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^m$, $m < n$, be a smooth map and let $\mathcal{X}_*(t)$ be as above, i.e. $F(x,t) = 0$ for $x \in \mathcal{X}_*(t)$, $t \in \mathbb{R}$. Assume further the existence of constants $R, L_1, L_2, L_3 > 0$ such that for all $t \in \mathbb{R}$ holds:*

1. $\|DF(x,t)\| \leq L_1$, *for all* $x \in \mathcal{X}_*(t)$,
2. $\sigma_{\min}(DF(x,t)^\top) > L_2$, *for all* $x \in \mathcal{X}_*(t)$,
3. $\|D^2F(x,t)\| \leq L_3$, *for all* $x \in U_R(t) := \{x \in \mathbb{R}^n | \operatorname{dist}(x, \mathcal{X}_*(t)) < R\}$.

*Then there exists* $0 < r < R$ *such that for any initial condition* $x(0)$ *with* $\operatorname{dist}(x(0), \mathcal{X}_*(0)) \leq r$ *there exists a unique solution* $x(t)$ *of*

$$\boxed{\dot{x} = DF(x,t)^\dagger \left(\mathcal{M}(x)F(x,t) - \tfrac{\partial}{\partial t}F(x,t)\right),} \tag{2.84}$$

*with the properties*

1. $\operatorname{dist}(x(t), \mathcal{X}_*(t)) < R$ *for all* $t \geq 0$.
2. $\operatorname{dist}(x(t), \mathcal{X}_*(t))$ *converges exponentially to* $0$.

*Proof.* Solutions $x(t)$ of (2.84) satisfy for $x(t) \in U_r(\mathcal{X}_*(t))$

$$\|F(x(t),t)\| \leq ae^{-bt},$$

for some $a, b, r > 0$, cf. (2.83). Therefore, these statements can be proven analogously to Main Theorem 2.1 by using the results of the previous lemma. □

**Underdetermined and inexact time-varying Newton flow**

Let $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^m$, $m < n$, as above and consider approximations $G(x,t)$ for $DF(x,t)^\dagger$, as described for $m = n$ in the previous section. Hence let $G(x,t) \in \mathbb{R}^{n\times m}$ such that the perturbation term

$$\Pi(x,t) := (DF(x,t)G(x,t) - I_m)\left(\mathcal{M}(x)F(x,t) - \frac{\partial}{\partial t}F(x,t)\right)$$

satisfies for some $\tilde{c}, R > 0$

$$\|\Pi(x,t)\| \leq \tilde{c}\|F(x,t)\|,$$

for all $x \in U_R(t) := \{x \in \mathbb{R}^n | \operatorname{dist}(x, \mathcal{X}_*(t)) < R\}$, $t \in \mathbb{R}$. We then approximate the ODE (2.84) by

$$\dot{x} = G(x,t)\left(\mathcal{M}(x)F(x,t) - \frac{\partial}{\partial t}F(x,t)\right). \tag{2.85}$$

It is straightforward to show, that the claims made in the previous theorem for equation (2.84) also hold qualitatively for (2.85), if $\mathcal{M}$ is a stable bundle map s.th. $\sup_{v\in\mathbb{R}^n, \|v\|=1} \langle \mathcal{M}(x) \cdot v, v\rangle < 0$ is sufficiently small. We therefore formulate the discretization of the underdetermined Newton flow such that it also includes the inexact case.

**Underdetermined and inexact time-varying Newton algorithm**
Let $t_k = kh$ for $h > 0$ and $k \in \mathbb{N}_0$. The discretization of equation (2.85) is given for $\mathcal{M}(x) = -\frac{1}{h}$ by

$$\boxed{x_{k+1} = x_k - G(x_k, t_k)\left(F(x_k, t_k) + hF_\tau^h(x_k, t_k)\right),} \tag{2.86}$$

where we used an approximation $F_\tau^h(x_k, t_k)$ of $\frac{\partial}{\partial t}F(x_k, t_k)$. We arrive at our most general tracking algorithm in Euclidean space.

**Main Theorem 2.4.** *Let $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^m$, $(x, t) \mapsto F(x, t)$ be smooth and let $\mathcal{X}_*(t)$ as above such that $F(x, t) = 0$ for all $x \in \mathcal{X}_*(t)$, $t \in \mathbb{R}$. Assume further the existence of constants $c_1, c_2, c_3, c_4, c_5, c_6, c_7, R > 0$ such that*

*(i)* $\|DF(x,t)\| \leq c_1$, $\|\frac{\partial}{\partial t}F(x,t)\| \leq c_2$, $\sigma_{\min}(DF(x,t)^\top) > c_3$ *for all* $x \in \mathcal{X}_*(t)$, $t \in \mathbb{R}$

*(ii)* $\|D^2F(x,t)\| \leq c_4$, $\|\frac{\partial^2}{\partial t^2}F(x,t)\| \leq c_5$, $\|\frac{\partial}{\partial t}DF(x,t)\| \leq c_6$ *for all* $x \in U_R(t) := \{x \in \mathbb{R}^n \mid \operatorname{dist}(x, \mathcal{X}_*(t)) < R\}$, $t \in \mathbb{R}$.

*(iii)* $\|F_\tau^h(x,t) - \frac{\partial F}{\partial t}(x,t)\| \leq c_7 h$, *for* $x \in U_R(t)$, $t \in \mathbb{R}$, $h > 0$.

*Let $G(x, t)$ denote an approximation for $DF(x, t)^\dagger$ satisfying for some $c, \tilde{c} > 0$*

$$\left\|(DF(x,t)G(x,t) - I_m)\left(\frac{1}{h}F(x,t) + F_\tau^h(x,t)\right)\right\| \leq \tilde{c}\|F(x,t)\|, \tag{2.87}$$

*for all $h > 0$ and $x \in U_{ch}(t)$, $t \in \mathbb{R}$.*
*Then the following statements hold*

1. *There exists $0 < r \leq R$ and $c_8, c_9, c_{10} > 0$ such that for $t \in \mathbb{R}$*

$$\operatorname{dist}(x, \mathcal{X}_*(t)) \leq c_8\|F(x,t)\|, \tag{2.88}$$

$$\|F(x,t)\| \leq c_9 \operatorname{dist}(x, \mathcal{X}_*(t)), \tag{2.89}$$

   *and*

$$\|DF(x,t)^\dagger\| \leq c_{10} \tag{2.90}$$

   *for $x \in U_r(t) := \{x \in \mathbb{R}^n \mid \operatorname{dist}(x, \mathcal{X}_*(t)) < r\}$.*

2. *The discretization sequence $(x_k)$ as defined in (2.86) with $t_k = kh$, $h > 0$ satisfies for some $c_{11}, c_{12} > 0$*

$$\operatorname{dist}(x_{k+1}, \mathcal{X}_*(t_{k+1})) \leq c_{11}\operatorname{dist}(x_k, \mathcal{X}_*(t_k))^2 + c_{12}h^2 \tag{2.91}$$

   *for $x_k \in U_r(t)$ with $r \leq ch$, $k \in \mathbb{N}_0$.*

3. *Let $c > 0$ be constant and $h$ sufficiently small. For any initial condition $x_0$ with $\|x_0 - x_*(0)\| < ch$ we have*

$$\text{dist}(x_k, \mathcal{X}_*(t_k)) \leq ch$$

*for all $k \in \mathbb{N}_0$. Thus, the update scheme* (2.86) *is well defined and produces estimates for $x_*(t_k)$, whose accuracy can be controlled by the step size.*

*Proof.* The first claim has been shown in Lemma 2.7. Therefore, claim 2) and 3) can be shown analogously to the proof of the second and third claim of Theorem 2.12. □

# Chapter 3

# Application I: Intrinsic Subspace Tracking

In this chapter we apply the general tracking techniques of the previous chapter to derive iterative algorithms for computing the principal subspace of time-varying symmetric matrices. The principal subspace of a symmetric matrix $A \in \mathbb{R}^{n\times n}$ is the $m$-dimensional eigenspace $V$, corresponding to the $m$ largest eigenvalues of $A$ for some $1 \leq m \leq n$.
The proposed algorithms are defined directly on the manifold; i.e. they are constructed either in a coordinate free way or via local coordinates. No Lagrange multiplier techniques are used or needed, nor any projection techniques that attempt to find solutions by projecting back suitable ambient space approximations. For this reason we refer to our algorithms as intrinsic. Our approach is motivated by [33], where new intrinsic implementations of the Newton method on Grassmann manifolds are introduced; thus improving earlier constructions by [23]. This approach can be easily extended to the time-varying problems, as is shown here. This leads us to particularly simple update schemes which robustly perform the tracking task.

## 3.1 Time-varying Newton flow for principal subspace tracking

We now consider the task of determining the non-constant $m$-dimensional principal subspace $V(t)$ of a family of symmetic matrices $A(t) \in \mathbb{R}^{n\times n}$, $t \in \mathbb{R}$, with eigenvalues $\lambda_1(t) \geq ... \geq \lambda_n(t)$. In order to derive the subspace tracking algorithms, we make the following assumptions:

**A1** The map $t \mapsto A(t) \in \mathrm{Sym}(n)$ is $C^r$ for some integer $r \geq 2$ and all $t \in \mathbb{R}$.

**A2** $\|A(t)\|$, $\|\dot{A}(t)\|$ and $\|\ddot{A}(t)\|$ are uniformly bounded for $t \in \mathbb{R}$.

**A3** The $m$ largest eigenvalues of $A(t)$ are well separated from the others, i.e. there exists $c > 0$ such that
$$\lambda_i(t) - \lambda_j(t) \geq c,$$

for $1 \leq i \leq m$ and $m+1 \leq j \leq n$, $t \in \mathbb{R}$.

Note that conditions (A1) and (A3) imply the existence of a $C^r$-curve $t \mapsto R(t) \in O(n)$ of orthogonal transformations such that

$$R(t)^\top A(t) R(t) = \operatorname{diag}(D_1(t), D_2(t)),$$

where $D_1(t) \in \mathrm{Sym}(m)$ has the eigenvalues $\lambda_1(t), ..., \lambda_m(t)$ and $D_2(t) \in \mathrm{Sym}(n-m)$ has the eigenvalues $\lambda_{m+1}(t), ..., \lambda_n(t)$ cf. Dieci and Eirola [20]. Note, that the blocks $D_1, D_2$ are not assumed to be diagonal. Thus for $[R_1(t)\ R_2(t)] := R(t)$ with $R_1(t) \in \mathbb{R}^{n\times m}$ we get $R_1(t)^\top R_1(t) = I_m$ and

$$R_1(t)^\top A(t) R_1(t) = D_1(t).$$

The span of the columns of $R_1(t)$ is therefore the principal subspace of $A(t)$, i.e. $V(t) = Im(R_1(t))$. Note that for any $Q \in O(m)$, the columns of $R_1(t)Q$ also span the principal subspace of $A(t)$.
To eliminate this ambiguity, we work on the Grassmann manifold, or more conveniently, on the Grassmannian $\mathrm{Gr}_{m,n}$. Recall, that the Grassmann manifold $\mathrm{Grass}(m,n)$ is defined as the set of $m$-dimensional subspaces in $\mathbb{R}^n$. It is well known that $\mathrm{Grass}(m,n)$ is a compact smooth manifold of dimension $m(n-m)$. In the sequel we prefer to work with an alternative, equivalent definition of the Grassmann manifold via the so-called *Grassmannian* $\mathrm{Gr}_{m,n}$, which is defined as the set of rank $m$ selfadjoint projection operators in $\mathbb{R}^n$:

$$\mathrm{Gr}_{m,n} := \{P \in \mathbb{R}^{n\times n} \mid P^\top = P, P^2 = P, \operatorname{tr} P = m\}. \tag{3.1}$$

It is well known, cf. Helmke and Moore [35], that $f : \mathrm{Gr}_{m,n} \to \mathrm{Grass}(m,n)$, $P \mapsto Im(P)$ defines a smooth diffeomorphism. Note, that the Grassmannian $\mathrm{Gr}_{m,n}$ is a Riemannian submanifold of the vector space of all symmetric matrices $\mathrm{Sym}_n := \{S \in \mathbb{R}^{n\times n} \mid S^\top = S\}$, endowed with the Frobenius inner product. Thus the Frobenius inner product now assumes the role of the Euclidean inner product in $\mathbb{R}^n$ and defines a Riemannian metric

$$< \xi, \eta >:= \operatorname{tr}(\xi\eta)$$

on each tangent space $T_P\mathrm{Gr}_{m,n}$, which is given by

$$T_P\mathrm{Gr}_{m,n} = \{[P, \Omega] \mid \Omega \in \mathfrak{so}_\mathrm{n}\}. \tag{3.2}$$

The information about the principal subspaces of a time-varying family of symmetric matrices $A(t)$ is now fully stored in the $C^r$-family of projection operators $P_*(t) := R_1(t)R_1(t)^\top \in \mathrm{Gr}_{m,n}$. Thus, in order to characterize the dominant subspaces of $A(t)$, we consider the trace function (Rayleigh quotient function) $f : \mathrm{Gr}_{m,n} \times \mathbb{R} \to \mathbb{R}$, defined by

$$f(P,t) = \operatorname{tr}(A(t)P). \tag{3.3}$$

It is well known that the maximum value of $\operatorname{tr}(A(t)P)$ is $\sum_{i=1}^m \lambda_i(t)$, implying that $P_*$ maximizes $f(P,t)$ if and only if $P_* = R_1(t)R_1(t)^\top$. Hence, the task of tracking the principal subspace of $A(t)$ is equivalent to track the points maximizing $f(P,t)$. For this reason, we can characterize the invariant subspaces of $A(t)$ via the zeros of the time-varying gradient vector field $\operatorname{grad} f(P,t)$. From [33, 63], explicit forms for the *Riemannian gradient* $\operatorname{grad} f$ and the *Riemannian Hessian operator* of the Rayleigh quotient function (3.3) are

$$\begin{aligned} \operatorname{grad} f(P,t) &= [P,[P,A(t)]], && (3.4)\\ H_f(P,t)\cdot X &= [P,[X,A(t)]], \quad X \in T_P\mathrm{Gr}_{m,n} && (3.5) \end{aligned}$$

The next lemma shows that the points of the curve $P_*(t)$, which maximize $f(P,t)$, are well separated from other critical points.

**Lemma 3.1.** *Let $A(t)$ satisfy (A1)-(A3). Then for $t \in \mathbb{R}$ and $P$ sufficiently close to $P_*(t)$, the condition* $\operatorname{grad} f(P,t) = 0$ *holds if and only if $P = P_*(t)$.*

*Proof.* The result follows by assumption on $P$, since $Im(P_*(t))$ is an isolated invariant subspace of $A(t)$.

□

We next derive bounds for the norm of the Riemannian Hessian.

**Lemma 3.2.** *Let $A(t)$ satisfy (A1)-(A3). Then there exist constants $M_1, M_2 > 0$ such that the Hessian $H_f$ satisfies*

$$M_1 \leq \|H_f(P_*(t),t)\cdot\xi\| \leq M_2,$$

*for all $t \in \mathbb{R}$ and all $\xi \in T_{P_*(t)}\mathrm{Gr}_{m,n}$ with $\|\xi\| = 1$.*

*Proof.* Here and in the following, substitute the symbol $\hat{I}$ to the matrix $\begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix}$. Let $P_*(t) =: \Theta\hat{I}\Theta^\top$, $\Omega \in \mathfrak{so}(\mathrm{n})$ and $\xi = [P_*(t),\Omega]$. Consider

$$\Theta^\top(H_f(P_*(t),t)\cdot\xi)\Theta = -\Theta^\top[P_*(t),[A,[P_*(t),\Omega]]]\Theta = -[\hat{I},[\Theta^\top A\Theta,[\hat{I},\Theta^\top\Omega\Theta]]].$$

Note that $Z := \Theta^\top\Omega\Theta \in \mathfrak{so}(\mathrm{n})$ satisfies for $Z = \begin{bmatrix} Z_1 & Z_2 \\ -Z_2^\top & Z_3 \end{bmatrix}$

$$[\hat{I},Z] = \begin{bmatrix} 0 & Z_2 \\ Z_2^\top & 0 \end{bmatrix} \in \mathrm{Sym}(n).$$

Thus,

$$\Theta^\top(H_f(P_*(t),t)\cdot\xi)\Theta = -[\hat{I},[N,\begin{bmatrix} 0 & Z_2 \\ Z_2^\top & 0 \end{bmatrix}]],$$

where $N := \Theta^\top A\Theta \in \mathrm{Sym}(n)$. By using $N = \begin{bmatrix} N_1 & N_2 \\ N_2^\top & N_3 \end{bmatrix}$, we obtain

$$\Theta^\top(H_f(P_*(t),t)\cdot\xi)\Theta = \begin{bmatrix} 0 & Z_2N_3 - N_1Z_2 \\ N_3Z_2^\top - Z_2^\top N_1 & 0 \end{bmatrix},$$

showing that

$$\|(H_f(P_*(t),t)\cdot\xi)\| = \|\Theta^\top(H_f(P_*(t),t)\cdot\xi)\Theta\| = \sqrt{2}\|Z_2N_3 - N_1Z_2\|. \tag{3.6}$$

Note that $N_1 = \Theta_1^\top A(t)\Theta_1$ and $N_3 = \Theta_2^\top A(t)\Theta_2$, where the columns of $\Theta_1$ span the principal subspace of $A(t)$, while $\Theta_2$ spans the complementary subspace of $A(t)$. Thus

$$\|Z_2N_3 - N_1Z_2\| \geq \min_{\substack{1\leq i\leq m\\ m+1\leq j\leq n}} (\lambda_i(t)-\lambda_j(t))\,\|Z_2\|,$$

which implies that

$$\|Z_2N_3 - N_1Z_2\| \geq c\|Z_2\|,$$

for some $c>0$, due to assumption (A3). Note further that $\xi = [P_*(t),\Omega] \in T_{P_*(t)}\mathrm{Gr}_{m,n}$ satisfies

$$\Theta^\top\xi\Theta = [\hat{I}, Z] = \begin{bmatrix} 0 & Z_2 \\ Z_2^\top & 0 \end{bmatrix},$$

which shows, that $\|Z_2\| = \frac{\|\xi\|}{\sqrt{2}}$. Thus,

$$\|(H_f(P_*(t),t)\cdot\xi)\| \geq c\;\|\xi\|.$$

On the other hand, we get from (3.6), that

$$\|(H_f(P_*(t),t)\cdot\xi)\| \leq \sqrt{2}\|Z_2\| \max_{\substack{1\leq i\leq m\\ m+1\leq j\leq n}} (\lambda_i(t)-\lambda_j(t))$$

$$\leq \|\xi\| \max_{\substack{1\leq i\leq m\\ m+1\leq j\leq n}} (\lambda_i(t)-\lambda_j(t)) \leq 2\|\xi\|\|A(t)\|.$$

The result follows, using assumption (A2). □

We are now able to compute the time-varying Newton flow for the special situation here. Hence, by using (3.4) and (3.5), the differential equation (2.33) becomes the implicit differential equation

$$\boxed{[P,[\dot{P},A(t)]] = -[P,[P,\dot{A}(t)]] - \sigma[P,[P,A(t)]].} \tag{3.7}$$

Thus, by applying Theorem 2.3 to this flow, we arrive at the following convergence result for the solutions of (3.7) .

**Theorem 3.1.** *Let $A(t)$ satisfy assumptions (A1)-(A3). Then the solution $P(t)$ of (3.7) exists for all $t>0$ and converges exponentially to $P_*(t)$, provided that $P(0)$ is chosen sufficiently close to $P_*(0)$.*

*Proof.* All we have to check is, whether the function (3.3) satisfies the conditions (1), (2) of Theorem 2.3. This is easily done, using the above Lemma. Since $\mathrm{Gr}_{m,n}$ is a compact Riemannian submanifold of $\mathbb{R}^n$, $\mathrm{Gr}_{m,n}$ is complete with $i^*(\mathrm{Gr}_{m,n}) > 0$. Condition (1) follows directly from Lemma 3.2, since $H_f(P,t)\cdot\xi = \pi_P D\hat{F}(P,t)\cdot\xi$ for $\hat{F}(P,t) = \operatorname{grad} f(P,t)$. Here, $\pi_P : \mathrm{Sym}_n \to T_P\mathrm{Gr}_{m,n}$ denotes the orthogonal projection onto the tangent space.
To verify Condition (2), consider the derivative of $H_f$, which is given by

$$\pi_P D(H_f(P,t)\cdot\xi)\cdot\eta = -\pi_P[\eta,[A(t),\xi]],$$

where $\xi,\eta \in T_P\mathrm{Gr}_{m,n}$. Thus for $\|\xi\| = \|\eta\| = 1$, we get that

$$\|\pi_P D(H_f(P,t)\cdot\xi)\cdot\eta\| \leq 4\|A(t)\|,$$

which completes the proof, since assumption (A2) holds. $\square$

## 3.2 Subspace tracking algorithms

The above differential equation for subspace tracking is implicit and thus hard to solve numerically. In this section, we therefore abandon the idea of working directly with the continuous time flow (3.7) and focus instead on suitable discretized algorithms. Thus, we specify the general time-varying Newton and parameterized time-varying Newton algorithm of Section 2 to the situation at hand. This leads us to explicit new numerical algorithms (Algorithms 2-4) for subspace tracking. Algorithm 1 is well-known from the work of [23], [44] and provides the benchmark for our subsequent algorithms. It is implemented by iterative solutions to matrix Sylvester equations, in conjunction with iterative computations of geodesics of the Grassmannian via singular value decompositions. However, no time adaptation step is made, to compensate for the time dependency of the vector fields. Algorithms 2-4 employ solutions to matrix Sylvester equations for smaller scale matrices; Algorithms 3,4 also incorporate adaptive terms to reflect time dependency effects. Instead of working with exact formulas for the geodesics in Algorithms 2,4, we find it more convenient to use approximations of the matrix exponential via the $QR$-factorization. This has the advantage that the $QR$-factorization can be exactly computed in finitely many steps, while all algorithms for computing matrix exponentials or singular value decompositions are inherently iterative.

**Algorithm 1: Riemannian Newton**
This method uses the standard Riemannian Newton method to track the time-varying extremum $x_*(t)$ of a cost function $f : \mathrm{Grass}(m,n)\times\mathbb{R} \to \mathbb{R}$ on the Grassmann manifold and is due to Lundström and Elden [44]. The authors did not give explicit formulas, but mentioned, that the standard Riemannian Newton algorithm was used to compute the approximation $x_{k+1}$ of $x_*(t_{k+1})$. Hence, by using the notation of Section 2, the update rule is given by

$$x_{k+1} = \exp_{x_k}\left(-H_f(x_k,t_{k+1})^{-1}\operatorname{grad} f(x_k,t_{k+1})\right). \tag{3.8}$$

Note that the use of $t_{k+1}$ in this update rule is crucial to perform a Newton update step towards $x_*(t_{k+1})$: If we used $t_k$ instead, then the step would go towards $t_*(t_k)$ and would lead qualitatively to the same tracking algorithm (up to renumbering). If we used both, i.e. $t_k$ in the Hessian and $t_{k+1}$ in the gradient or vice versa, we would not perform a real Newton update step but an inexact Newton method.
Thus the only possible implementation is given by formula (3.8), which obviously works, if all iterates $x_k$ lie in the domain of attraction of the Newton method of grad $f(\cdot, t_{k+1})$ for $k \in \mathbb{N}$. The implementation of the algorithm in [44] then follows the computations in Edelman, Arias and Smith [23]. However, the convergence properties of (3.8) for time-varying problems have not been investigated in [44], nor in any previous publication we are aware of.
Thus the Riemannian Newton tracking algorithm is described using matrices $Y(t) \in \mathbb{R}^{n\times m}$ satisfying $Y(t)^\top Y(t) = I_m$, where the principal subspace of $A(t)$ is given by $Im(Y(t))$. To compute $Y(t)$ and the corresponding point $P(t)$ on the Grassmannian at discrete times $t_k = kh$ for $k \in \mathbb{N}$ and $h > 0$, the task is to maximize the function $f(Y,t) := \frac{1}{2}\mathrm{tr}\,(Y^\top A(t) Y)$.
As it turned out, this leads to the computation of the solution $\Delta_k$ of the Sylvester equation

$$\Pi_k A(t_{k+1})\Pi_k \Delta_k - \Delta_k H_k = -R_k,$$

where $\Pi_k = I - Y_k Y_k^\top$, $H_k = Y_k^\top A(t_{k+1}) Y_k$ and $R_k = A(t_{k+1})Y_k - Y_k H_k$. Then the sequences $(Y_k)$ and $(P_k)$ are defined by

$$\boxed{\begin{array}{c} Y_{k+1} = Y_k V_k \cos(\Sigma_k) V_k^\top + U_k \sin(\Sigma_k) V_k^\top \\ P_{k+1} = Y_{k+1} Y_{k+1}^\top, \end{array}} \tag{3.9}$$

where $U_k \Sigma_k V_k^\top$ is the compact singular value decomposition of $\Delta_k$. Thus, this algorithm requires at each step to solve a Sylvester equation for $\Delta \in \mathbb{R}^{n\times m}$ and the computation of the SVD of $\Delta$. Therefore, this method uses an overparameterization of the Grassmannian by the Stiefel manifold.
More efficient implementations of the Riemannian Newton algorithm are possible using parameterized versions of the Newton method as in [33], which in turn is based on the earlier work of [61], [38] and [47]. Since these implementations will also appear in our subsequent algorithms we do not give a parallel treatment here.

**Algorithm 2: Parameterized Riemannian Newton**
We now consider a subspace tracking method, which originally was derived for optimization on the Grassmann manifold in [33]. In order to specify this approach for maximization of the time-varying cost function $f : \mathrm{Gr}_{m,n} \times \mathbb{R} \to \mathbb{R}$, defined by $P \mapsto \mathrm{tr}\,(A(t)P)$, we consider arbitrary families of smooth local uniform parameterizations $\gamma_P : T_P\mathrm{Gr}_{m,n} \to \mathrm{Gr}_{m,n}$ and $\mu_P : T_P\mathrm{Gr}_{m,n} \to \mathrm{Gr}_{m,n}$ of the Grassmannian for $P \in \mathrm{Gr}_{m,n}$. The proposed *parameterized Riemannian Newton update scheme* then is

$$P_{k+1} = \mu_{P_k}\left(-H_{f\circ\hat{\gamma}_{P_k}}(0,t_{k+1})^{-1}\nabla(f\circ\hat{\gamma}_{P_k})(0,t_{k+1})\right), \tag{3.10}$$

where $\hat{\gamma}_P(v,t) := (\gamma_P(v),t)$ for $P \in \mathrm{Gr}_{m,n}$, $v \in T_P\mathrm{Gr}_{m,n}$ and $t \in \mathbb{R}$. Note, that if both families of parameterizations $\mu_P$ and $\gamma_P$ are chosen to be equal to the Riemannian exponential map $\exp_P : T_P\mathrm{Gr}_{m,n} \to \mathrm{Gr}_{m,n}$, then (3.10) becomes equivalent to the Riemannian Newton method (3.8). Other choices of local parameterizations therefore lead to modifications of the Riemannian Newton method. In the sequel, we find it convenient to replace the Riemannian exponential map by *QR-coordinates on the Grassmannian*. In contrast to this, the parameterization $\gamma_P$ is set to the Riemannian exponential map $\exp_P$. Then, the parameterized gradient and parameterized Hessian are equal to the Riemannian gradient and Hessian, cf. [33] for details. Thus, (3.10) turns into

$$P_{k+1} = \mu_{P_k}\left(-H_f(P_k,t_{k+1})^{-1}\mathrm{grad}\, f(P_k,t_{k+1})\right), \tag{3.11}$$

which leads to the following tracking algorithm for the principal subspace $P_k$ of a time-varying symmetric matrix $A(t)$ at discrete times $t_k = kh$ for $k \in \mathbb{N}$ and $h > 0$. Let $P_0 \in \mathrm{Gr}_{m,n}$ with $\Theta_0 \in O(n)$ such that

$$P_0 = \Theta_0 \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta_0^\top.$$

Then, assuming that $\Theta_k$ has been computed already, we define

$$\begin{bmatrix} N_1 & N_2 \\ N_2^\top & N_3 \end{bmatrix} := \Theta_k^\top A(t_{k+1})\Theta_k.$$

Let $Z_k \in \mathbb{R}^{m\times(n-m)}$ denote the solution to the Sylvester equation

$$N_1 Z_k - Z_k N_3 = N_2.$$

Then the next iteration step is given as

$$P_{k+1} = \Theta_{k+1} \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta_{k+1}^\top \tag{3.12}$$

where

$$\boxed{\Theta_{k+1} = \Theta_k \begin{bmatrix} I_m & -Z_k \\ Z_k^\top & I_{n-m} \end{bmatrix}_{\mathrm{Q}}}$$

Hence to use this algorithm, one needs at each step to solve a Sylvester equation for $Z \in \mathbb{R}^{m\times(n-m)}$ and the execution of the QR-algorithm for a $m \times n$ and a $(n-m)\times n$ matrix, cf. Remark 3.1. Therefore, this algorithm has computationally advantages, compared to Algorithm 1, which requires that a Sylvester equation is solved on matrix space of dimension $mn$. In contrast, the Sylvester equation in Algorithm 2 has to be solved on a matrix space of dimension $m(n-m)$. Since the dimension of the Grassmann manifold is $m(n-m)$, this is the minimal number of parameters that have to be computed. This property of working with the minimal number of parameters is also true for the subsequent algorithms. The resulting advantages are confirmed by numerical examples in the next section.

**Remark 3.1.** The computation of $\begin{bmatrix} I_m & -Z_k \\ Z_k^\top & I_{n-m} \end{bmatrix}_{\mathrm{Q}}$ can be effectively done by computing the $Q$-factor of the block columns separately, i.e.

$$\begin{bmatrix} I_m & -Z_k \\ Z_k^\top & I_{n-m} \end{bmatrix}_{\mathrm{Q}} = \left( \begin{bmatrix} I_m \\ Z_k^\top \end{bmatrix}_{\mathrm{Q}} \begin{bmatrix} -Z_k \\ I_{n-m} \end{bmatrix}_{\mathrm{Q}} \right).$$

This relation holds since the first block column $[I_m \ Z_k]^\top$ is orthogonal to $[-Z_k^\top \ I_{n-m}]^\top$.

**Algorithm 3: Time-varying Newton**
In contrast to the previous two methods, this method is derived from a truly time-varying approach. It comes from the discretization of the Newton flow (3.7) and computes approximations $P_k$ of the principal subspace $P_*(t_k)$ of $A(t_k)$ at discrete times $t_k = kh$ for $k \in \mathbb{N}$ and step size $h > 0$. It is given by

$$P_{k+1} = \exp_{P_k}\left(-H_f(P_k, t_k)^{-1}\left(\operatorname{grad} f(P_k, t_k) + hG_h(P_k, t_k)\right)\right), \tag{3.13}$$

where $\exp_{P_k}$ denotes the exponential map of $\mathrm{Gr}_{m,n}$ at $P_k$ and $G_h(P,t)$ is a suitable approximation of $\frac{\partial}{\partial t}\operatorname{grad} f(P,t)$. Thus, by using an approximation $A_h(t)$ of $\dot{A}(t)$, we set $G_h(P_k, t_k) = [P_k, [P_k, A_h(t_k)]]$.
To implement the discrete tracking algorithm (3.13) however, we need formulas for the inverse Hessian operator and the exponential map. Note, that for $\xi = [P, \Omega]$ ($\Omega \in \mathfrak{so}(\mathrm{n})$) the equation

$$H_f(P,t)\xi = -\operatorname{grad} f(P,t) - hG_h(P,t),$$

becomes equivalent to

$$[P, [A(t), [P, \Omega]]] = [P, [P, A(t) + hA_h(t)]],$$

and thus to

$$\Theta^\top [P, [A(t), [P, \Omega]]]\Theta = \Theta^\top [P, [P, A(t) + hA_h(t)]]\Theta, \qquad \text{where} \ \ \Theta \in \mathrm{O(n)}.$$

Thus, for $P =: \Theta \hat{I} \Theta^\top$, the above equation turns into

$$[\hat{I}, [\Theta^\top A(t)\Theta, [\hat{I}, \Theta^\top \Omega \Theta]]] = [\hat{I}, [\hat{I}, \Theta^\top (A(t) + hA_h(t))\Theta]]. \tag{3.14}$$

Let $N := \Theta^\top A(t)\Theta \in \mathrm{Sym}(n)$, $N(t) = \begin{bmatrix} N_1 & N_2 \\ N_2^\top & N_3 \end{bmatrix}$ and let $M := \Theta^\top (A(t) + hA_h(t))\Theta \in \mathrm{Sym}(n)$, $M(t) = \begin{bmatrix} M_1 & M_2 \\ M_2^\top & M_3 \end{bmatrix}$. Note that $Z := \Theta \Omega \Theta^\top \in \mathfrak{so}(\mathrm{n})$, $Z = \begin{bmatrix} Z_1 & Z_2 \\ -Z_2^\top & Z_3 \end{bmatrix}$ satisfies

$$[\hat{I}, Z] = \begin{bmatrix} 0 & Z_2 \\ Z_2^\top & 0 \end{bmatrix} \in \mathrm{Sym}(n).$$

Thus, the equation (3.14) is equivalent to

$$[\hat{I}, [N, \begin{bmatrix} 0 & Z_2 \\ Z_2^\top & 0 \end{bmatrix}]] = [\hat{I}, [\hat{I}, M]],$$

which shows, that the inverse of $H_f$ is obtained by solving the following Sylvester equation for $Z_2$:

$$N_1 Z_2 - Z_2 N_3 = M_2 \tag{3.15}$$

This is the first result needed for the implementation of the discrete update scheme. We also need the formula of the exponential map of $\mathrm{Gr}_{m,n}$, which is given at $P = \Theta \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta^\top$, $\Theta \in O(n)$, by

$$\exp_P(\xi) = \Theta \begin{bmatrix} \cos\sqrt{Z_2 Z_2^\top} \\ -\frac{\sin\sqrt{Z_2^\top Z_2}}{\sqrt{Z_2^\top Z_2}} Z_2^\top \end{bmatrix} \begin{bmatrix} \cos\sqrt{Z_2 Z_2^\top} & -Z_2 \frac{\sin\sqrt{Z_2^\top Z_2}}{\sqrt{Z_2^\top Z_2}} \end{bmatrix} \Theta^\top, \tag{3.16}$$

cf. [33]. Here, $Z_2$ is also defined for $\xi = [P, \Omega] \in T_P\mathrm{Gr}_{m,n}$ by $Z := \Theta^\top \Omega \Theta \in \mathfrak{so}(\mathrm{n})$ with $Z = \begin{bmatrix} Z_1 & Z_2 \\ -Z_2^\top & Z_3 \end{bmatrix}$.

Thus, the update scheme (3.13) can be rewritten as

$$\boxed{P_{k+1} = \Theta_k \begin{bmatrix} \cos\sqrt{Z_2 Z_2^\top} \\ -\frac{\sin\sqrt{Z_2^\top Z_2}}{\sqrt{Z_2^\top Z_2}} Z_2^\top \end{bmatrix} \begin{bmatrix} \cos\sqrt{Z_2 Z_2^\top} & -Z_2 \frac{\sin\sqrt{Z_2^\top Z_2}}{\sqrt{Z_2^\top Z_2}} \end{bmatrix} \Theta_k^\top,} \tag{3.17}$$

where $Z_2$ solves the Sylvester equation (3.15) and $\Theta_k \in O(n)$ satisfies

$$P_k = \Theta_k \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta_k^\top.$$

The convergence property of this update scheme is characterized in the next theorem.

**Theorem 3.2.** *Let $A(t)$ satisfy (A1)-(A3). Let $A_h(t) = A_h(t)^\top$, $t \in \mathbb{R}$, be an arbitrary family of symmetric matrices such that $\|\dot{A}(t) - A_h(t)\| < \tilde{c}h$ holds for all $h > 0$, $t \in \mathbb{R}$ and some constant $\tilde{c} > 0$. Then for $c > 0$ and sufficiently small $h > 0$, the update scheme* (3.17) *satisfies for $k \in \mathbb{N}$*

$$\mathrm{dist}(P_k, P_*(t_k)) \leq ch,$$

*provided that* $\mathrm{dist}(P_0, P_*(0)) \leq ch$.

*Proof.* With $f$ defined as in (3.3), we obviously have that $G_h(P,t) := [P, [P, A_h(t)]]$ is an approximation of $\frac{\partial}{\partial t}\mathrm{grad}\, f(P,t)$ such that for some $R, \tilde{C} > 0$

$$\left\| G_h(P,t) - \frac{\partial}{\partial t}\mathrm{grad}\, f(P,t) \right\| \leq \tilde{C}h,$$

for all $h > 0$, $P \in \mathcal{B}_R(P_*(t))$, $t \in \mathbb{R}$. Moreover, it has been already shown in the proof of Theorem 3.1, that the assumptions of Theorem 2.3 are satisfied under these conditions. Note that condition 1 of Theorem 2.3 implies in particular, that $\|\pi_{T_{x_*(t)}M} D\hat{F}(x_*(t), t)^{-1}\|$ is bounded for $t \in \mathbb{R}$, where we used the notation of Theorem 2.3.
It remains only to establish bounds on the norm of some partial derivatives of the gradient that appear in Theorem 2.4. But this follows immediately from the identities

$$\frac{\partial}{\partial t}\operatorname{grad} f(P, t) = [P, [P, \dot{A}(t)]]$$

and

$$\frac{\partial^2}{\partial t^2}\operatorname{grad} f(P, t) = [P, [P, \ddot{A}(t)]],$$

together with the uniform boundedness of $\|P\|$, $\|\dot{A}(t)\|$ and $\|\ddot{A}(t)\|$ for $P \in \mathrm{Gr}_{m,n}$ and $t \in \mathbb{R}$. Thus the assumptions of Theorem 2.4 are satisfied and the result follows from Theorem 2.4. □

The update scheme (3.17) can be implemented as follows.

**Implementation of Algorithm 3 (Time-varying Newton)**

1. Choose the step size $h > 0$ and $P_0 \approx P_*(0)$ with $P_0 \in \mathrm{Gr}_{m,n}$ and set $k = 0$.
2. Pick an orthogonal matrix $\Theta_k \in O(n)$ such that
$$P_k = \Theta_k \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta_k^\top.$$
3. Compute for $t_k = kh$
$$\begin{bmatrix} N_1 & N_2 \\ N_2^\top & N_3 \end{bmatrix} = \Theta_k^\top A(t_k)\Theta_k$$
and
$$\begin{bmatrix} M_1 & M_2 \\ M_2^\top & M_3 \end{bmatrix} = \Theta_k^\top (A(t_k) + hA_h(t_k))\Theta_k.$$
4. Solve the Sylvester equation
$$N_1 Z_k - Z_k N_3 = M_2.$$
for $Z_k \in \mathbb{R}^{m\times(n-m)}$.
5. Compute
$$P_{k+1} = \Theta_k \begin{bmatrix} \cos\sqrt{Z_k Z_k^\top} \\ -\frac{\sin\sqrt{Z_k^\top Z_k}}{\sqrt{Z_k^\top Z_k}} Z_k^\top \end{bmatrix} \begin{bmatrix} \cos\sqrt{Z_k Z_k^\top} & -Z_k \frac{\sin\sqrt{Z_k^\top Z_k}}{\sqrt{Z_k^\top Z_k}} \end{bmatrix} \Theta_k^\top,$$
6. Set $k = k + 1$ and proceed with 2).

**Algorithm 4: Parameterized time-varying Newton**
Here we introduce a tracking algorithm for the time-varying principal subspace $P_*(t)$ of $A(t)$ by using parameterizations of the Grassmannian. It will turn out, that this approach leads to a much simpler update scheme than the formulas derived for Algorithm 3.
For $P \in \mathrm{Gr}_{m,n}$, we consider the family of smooth local parameterizations

$$\mu_P : T_P\mathrm{Gr}_{m,n} \to \mathrm{Gr}_{m,n},$$

$$\xi \mapsto (I + [\xi, P])_Q \, P \, (I + [\xi, P])_Q^\top,$$

where $(A)_Q$ denotes the $Q$ factor of the $QR$-factorization $A = QR = (A)_Q R$ of $A$. For this reason, we also call this parameterization QR-coordinates on the Grassmannian. It is easily seen, cf. [33], that the map $\mu_P$ is smooth on the tangent space $T_P\mathrm{Gr}_{m,n}$ with $\mu_P(0) = P$ and the derivative of the parameterization $D\mu_P(0) : T_P\mathrm{Gr}_{m,n} \to T_P\mathrm{Gr}_{m,n}$ is equal to the identity map id. Moreover,

$$\tfrac{\mathrm{d}^2}{\mathrm{d}\,\varepsilon^2}\mu_P(\varepsilon\xi)\Big|_{\varepsilon=0} = \Theta \begin{bmatrix} -2ZZ^\top & 0 \\ 0 & 0 \end{bmatrix} \Theta^\top, \tag{3.18}$$

where $\xi = \Theta \begin{bmatrix} 0 & Z \\ Z^\top & 0 \end{bmatrix} \Theta^\top$ for $P = \Theta \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta^\top$.
To generate a sequence $\{P_k\}$, which tracks the maximum $P_*(t)$ of $f$ at discrete times $t = t_k$ for $k \in \mathbb{N}$, $h > 0$, we consider the parameterized time-varying Newton algorithm

$$P_{k+1} = \mu_{P_k}\left(-H_f(P_k, t_k)^{-1}(\operatorname{grad} f(P_k, t_k) + hG_h(P_k, t_k))\right),$$

where $G_h(P_k, t)$ denotes an approximation of $\frac{\partial}{\partial t}\operatorname{grad} f(P_k, t)$, cf. equation (2.36). Note that we set the second parameterization $\gamma_P$ in (2.36) to be equal to the Riemannian normal coordinates, which enabled us to replace the parameterized gradient and Hessian of $f$ by the Riemannian gradient and Hessian in the above formula, cf. [33] for details. We now attempt to derive a more explicit form of this algorithm. Using an approximation $A_h(t)$ for $\dot{A}(t)$ and setting $G_h(P, t)$ to $[P, [P, A_h(t)]]$, the above equation is equivalent to

$$P_{k+1} = \mu_{P_k}\left((\operatorname{ad}_P \circ \operatorname{ad}_{A(t_k)})^{-1}\left([P, [P, A(t_k) + hA_h(t_k)]]\right)\right).$$

In the previous section, it has been shown, that $(\operatorname{ad}_P \circ \operatorname{ad}_{A(t)})^{-1}\left([P, [P, A(t) + hA_h(t)]]\right)$ equals $\Theta \begin{bmatrix} 0 & Z \\ Z^\top & 0 \end{bmatrix} \Theta^\top$, where $Z \in \mathbb{R}^{m\times(n-m)}$ solves equation (3.15). The update scheme therefore simplifies to

$$P_{k+1} = \mu_{P_k}\left(\Theta_k \begin{bmatrix} 0 & Z \\ Z^\top & 0 \end{bmatrix} \Theta_k^\top\right),$$

which can be conveniently rewritten as follows:

$$\boxed{P_{k+1} = \Theta_k \begin{bmatrix} I_m & -Z \\ Z^\top & I_{n-m} \end{bmatrix}_Q \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} I_m & -Z \\ Z^\top & I_{n-m} \end{bmatrix}_Q^\top \Theta_k^\top.} \tag{3.19}$$

The convergence properties of the algorithm are stated in the next result.

**Theorem 3.3.** *Under the same assumption as for Theorem 3.2, for $c > 0$ and sufficiently small $h > 0$, the sequence (3.19) satisfies*

$$\operatorname{dist}(P_k, P_*(t_k)) \le ch$$

*for all $k \in \mathbb{N}$ and $t_k = kh$, provided that $P_0$ is sufficiently close to $P_*(0)$.*

*Proof.* In order to apply Main Theorem 2.3, we check first the conditions regarding the parameterization $\mu_P(\xi) = (I + [\xi, P])_Q \, P \, (I + [\xi, P])_Q^\top$.
As mentioned before, the map $\mu_P$ is smooth on the tangent space $T_P\mathrm{Gr}_{m,n}$ with $\mu_P(0) = P$ and the derivative of the parameterization $D\mu_P(0) : T_P\mathrm{Gr}_{m,n} \to T_P\mathrm{Gr}_{m,n}$ is equal to the identity map id. Thus, it remains to check, if the $\|D^2\mu_P(\xi)\|$ is uniformly bounded for $\xi \in B_R(0)$, $P \in \mathrm{Gr}_{m,n}$ for some fixed $R > 0$.
For $P = \Theta \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta^\top$, and $\xi = \Theta \begin{bmatrix} 0 & Z \\ Z^\top & 0 \end{bmatrix} \Theta^\top$, we have $[\xi, P] = \Theta \begin{bmatrix} 0 & -Z \\ Z^\top & 0 \end{bmatrix} \Theta^\top$ and the parameterization turns into

$$\mu_P\left(\Theta \begin{bmatrix} 0 & Z \\ Z^\top & 0 \end{bmatrix} \Theta^\top\right) = \left(\Theta \begin{bmatrix} I & -Z \\ Z^\top & I \end{bmatrix} \Theta^\top\right)_Q P \left(\left(\Theta \begin{bmatrix} I & -Z \\ Z^\top & I \end{bmatrix} \Theta^\top\right)_Q\right)^\top,$$

which can be rewritten as

$$\mu_P\left(\Theta \begin{bmatrix} 0 & Z \\ Z^\top & 0 \end{bmatrix} \Theta^\top\right) = \Theta \begin{bmatrix} I & -Z \\ Z^\top & I \end{bmatrix}_Q \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \left(\begin{bmatrix} I & -Z \\ Z^\top & I \end{bmatrix}_Q\right)^\top \Theta^\top.$$

Now let $R : \mathbb{R}^{m\times(n-m)} \to O(n)$ be defined by $Z \mapsto \begin{bmatrix} I & -Z \\ Z^\top & I \end{bmatrix}_Q$ and note that $\begin{bmatrix} I & -Z \\ Z^\top & I \end{bmatrix}$ is invertible for all $Z \in \mathbb{R}^{m\times(n-m)}$. Since the QR factorization is smooth for general invertible matrices, cf. [33], we conclude, that the derivatives of $R$ exist and the norms of the derivatives $DR(Z) : \mathbb{R}^{m\times(n-m)} \to so(n)$ and $D^2R(Z) : \mathbb{R}^{m\times(n-m)} \times \mathbb{R}^{m\times(n-m)} \to so(n)$ are uniformly bounded for $Z \in B_R(0)$ for arbitrary but fixed $R > 0$. Since $\|\Theta\| = 1$, we get that $\|D^2\mu_P(\xi)\|$ is bounded for $\|\xi\| \le \sqrt{2}R$ and all $P \in \mathrm{Gr}_{m,n}$. This shows that the parameterization satisfies the necessary conditions.

With $f$ defined as in (3.3), we obviously have that $G_h(P, t) := [P, [P, A_h(t)]]$ is an approximation of $\frac{\partial}{\partial t}\mathrm{grad}\, f(P, t)$ such that for some $R, \tilde{C} > 0$

$$\left\| G_h(P, t) - \frac{\partial}{\partial t}\mathrm{grad}\, f(P, t) \right\| \le \tilde{C}h,$$

for all $h > 0$, $P \in \mathcal{B}_R(P_*(t))$, $t \in \mathbb{R}$.
Hence, to employ Main Theorem 2.3, it remains to show for some $\tilde{R} > 0$ the boundedness of

1. $\|H_f(P_*(t), t)\|^{-1}$ for all $t \in \mathbb{R}$,
2. $\|H_f(P_*(t), t)\|$ for all $t \in \mathbb{R}$,
3. $\|\frac{\partial}{\partial t}\mathrm{grad}\, f(P, t)\|$ for all $P \in \mathcal{B}_{\tilde{R}}(P_*(t))$, $t \in \mathbb{R}$,
4. $\|DH_f(P, t)\|$, $\|\frac{\partial^2}{\partial t^2}\mathrm{grad}\, f(P, t)\|$, $\|\frac{\partial}{\partial t}H_f(P, t)\|$ for all $P \in \mathcal{B}_{\tilde{R}}(P_*(t))$, $t \in \mathbb{R}$.

To see (1) it suffices to note Lemma 3.2. The second claim (2) follows from (3.5) and the boundedness of $\|A(t)\|$ and $\|P_*(t)\|$. The remaining conditions can be easily seen by computing the derivatives of (3.4) and (3.5), since $\|\dot{A}(t)\|$ and $\|\ddot{A}(t)\|$ are bounded. □

**Implementation of Algorithm 4 (Parameterized time-varying Newton)**
The update scheme (3.19) defines the following tracking algorithm.

1. Choose the step size $h > 0$, $P_0 \approx P_*(0)$ with $P_0 \in \mathrm{Gr}_{m,n}$ and pick an orthogonal matrix $\Theta_0 \in O(n)$ such that
$$P_0 = \Theta_0 \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta_0^\top,$$
and set $k = 0$.
2. Compute for $t_k = kh$
$$\begin{bmatrix} N_1 & N_2 \\ N_2^\top & N_3 \end{bmatrix} = \Theta_k^\top A(t_k)\Theta_k$$
and
$$\begin{bmatrix} M_1 & M_2 \\ M_2^\top & M_3 \end{bmatrix} = \Theta_k^\top (A(t_k) + hA_h(t_k))\Theta_k.$$
3. Solve the Sylvester equation
$$N_1 Z_k - Z_k N_3 = M_2$$
for $Z_k \in \mathbb{R}^{m\times(n-m)}$.
4. Compute
$$\Theta_{k+1} = \Theta_k \begin{bmatrix} I_m & -Z_k \\ Z_k^\top & I_{n-m} \end{bmatrix}_{\mathrm{Q}} \quad \text{and} \quad P_{k+1} = \Theta_{k+1} \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta_{k+1}^\top$$
5. Set $k = k + 1$ and proceed with 2).

Note that the implementation of this algorithm is considerably cheaper than the method defined by Theorem 3.2. In particular, the computation of $\Theta_{k+1}$ is much simpler than in Algorithm 3.

**Remark 3.2. Minor subspace tracking**
Computing the minor components of a matrix frequently arises in signal processing applications. Here the $m$-dimensional minor subspace of the symmetric matrix $A(t)$, is defined by the space spanned by the $m$ smallest eigenvectors of $A(t)$. Since the

algorithms for the principal subspace tracking are derived by maximizing the cost function $f(P,t) = \mathrm{tr}\,(A(t)P)$, it is trivial to use such algorithms for minor subspace analysis. In fact, we just have apply the above methods to the matrix $-A(t)$, since the principal subspace of $-A(t)$ is the minor subspace of $A(t)$. Therefore, the above algorithms can be used both for principal and minor component analysis.

## 3.3 Numerical results

All simulations were performed in Matlab, version 6.5. If not stated otherwise, we used $N = 40$ steps, step size $h = 0.025$, and considered the matrix

$$A(t) = X_*(t)K(t)X_*(t)^\top \tag{3.20}$$

for $K(t) = \mathrm{diag}\,(a_{10} + \sin(10t), ..., a_1 + \sin(1t))$,

$$X_*(t) = R^\top \begin{bmatrix} \cos(t) & \sin(t) & 0 \\ -\sin(t) & \cos(t) & 0 \\ 0 & 0 & I_8 \end{bmatrix} R. \tag{3.21}$$

Here, $R \in O(10)$ is a fixed random orthogonal matrix and $a_i := 2.5i$ for $i = 1, ..., 10$. We always used a 2nd order approximation $A_h(t)$ for $\dot{A}(t)$.

**Example 1: Investigation of the time-varying subspace tracking algorithms**
In the first simulation, we track the 3-dimensional principal subspace of $A(t_k)$ at discrete times $t_k = kh$ for $k = 1, ..., 40$. We applied the parameterized time-varying Newton method (Algorithm 4) to track $P_*(t_k)$, which significantly simplifies the computations, compared to Algorithm 3. In order to test the error correction of the method, we generate perturbed initial values $P' := P_*(0) + B$, where $B$ is a random matrix with entries in $(-0.2, 0.2)$. Then the perturbed starting point $P_0$ is obtained by orthogonalizing $P'$ via the QR-algorithm.
From the error plot, where $\|P_k - P_*(t_k)\|$ is shown, we observe a fast convergence to zero, cf. Figure 3.1, indicating the robustness of the chosen method. The mean error $\frac{1}{10}\sum_{k=31}^{40} \|P_k - P_*(t_k)\|$ of the 10 last points was about $1.2 \cdot 10^{-4}$.
In Figure 3.2, the trace of $A(t_k)P_k$ is compared to the sum of the first three principal eigenvalues $\lambda_1(t_k), ..., \lambda_3(t_k)$. We observe that after a few initial steps, these magnitudes are practically equal, confirming that $P_k$ approximatively maximizes $\mathrm{tr}\,(A(t_k)P)$.
Note that Algorithm 4 computes a matrix $\Theta_k$ and sets $P_k = \Theta_k \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta_k^\top$. This orthogonal matrix $\Theta_k$ has the property, that the leading columns $\Theta_k[I_3\ 0]^\top \in \mathbb{R}^{10\times 3}$ approximatively span the principal subspace of $A(t_k)$. We therefore computed the eigenvalues of $[I_3\ 0]\Theta_k^\top A(t_k)\Theta_k[I_3\ 0]^\top$ and compared them with the dominant eigenvalues of $A(t_k)$, cf. Figure 3.3. Here fast convergence of the eigenvalues of $[I_3\ 0]\Theta_k^\top A(t_k)\Theta_k[I_3\ 0]^\top$ to the principal eigenvalues of $A(t_k)$ is observed.
Next, we investigate the influence of the step size on the accuracy of Algorithm 4 by changing the step size after each run. We use perfect initial conditions, i.e. $P_0 = P_*(0)$

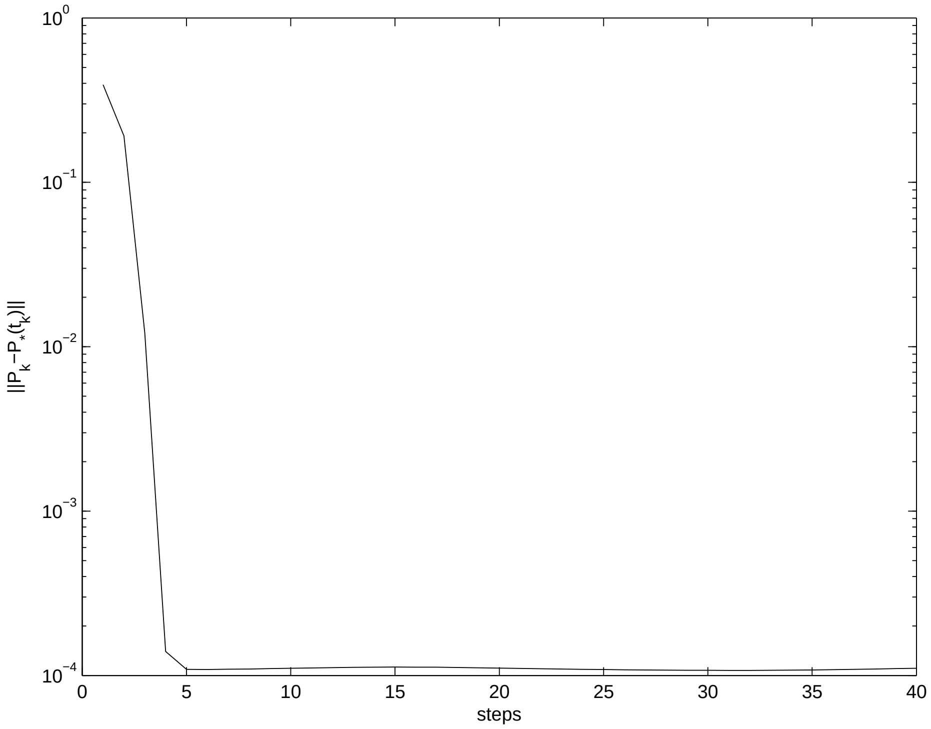


Figure 3.1: The evolution of the error $\|P_k - P_*(t_k)\|$. The parameterization method (3.19) was used.

| Step size h | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.1 |
|---|---|---|---|---|---|---|
| Accuracy | $2.1 \cdot 10^{-5}$ | $7.0 \cdot 10^{-5}$ | $1.4 \cdot 10^{-4}$ | $3.4 \cdot 10^{-4}$ | $4.7 \cdot 10^{-4}$ | $6.1 \cdot 10^{-3}$ |

Table 3.1: The accuracy of Algorithm 4 for different step sizes.

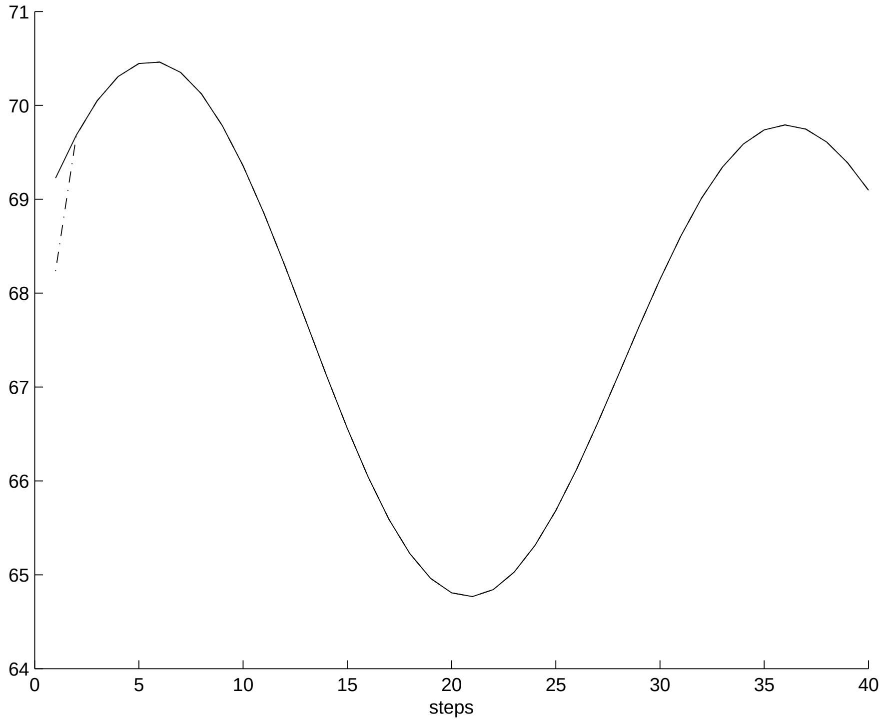


Figure 3.2: The evolution of tr $(P_k A(t_k))$ (dotted) compared to the sum of the principal eigenvalues $\lambda_1(t_k), \lambda_2(t_k), \lambda_3(t_k)$, corresponding to Figure 3.1.

and define the accuracy by $\frac{1}{40}\sum_{k=1}^{40} \|P_k - P_*(t_k)\|$. The results are given in Table 3.1 and show, that the error is increasing overproportionally to the step size.

**Example 2: Behavior near coalescing principal and minor eigenvalues**
Note that we needed to assume that the smallest principal eigenvalue is well separated from the minor eigenvalues, to prove the convergence of the tracking algorithms. Thus, we also checked the algorithm's behavior in the case of "almost coalescing principal and minor eigenvalues". Since the algorithm also works in case of some identical eigenvalues we considered the following setup:
The matrix $A(t)$ was defined as in (3.20), where now $K(t) = \mathrm{diag}\,(a_1(t), ..., a_{10}(t))$ for $a_1(t) = a_2(t) = a_3(t) = 2.001 - \sin(t)$ and $a_4(t) = ... = a_{10}(t) = \sin(t)$. We computed the 3-dimensional principal subspace $P_k \approx P_*(t_k)$ of $A(t_k)$ for $k = 1, ..., 80$ by using Algorithm 4. For $t = \pi/2$, we have $\lambda_3(t) = 1.001$ and $\lambda_4(t) = 1$. Thus for $k = 63$, we can expect a perturbation in the algorithm's output, since $t_{63} = 63 \times 0.025 = 1.575 \simeq \pi/2$. Figure 3.4 exhibits the error $\|P_k - P_*(t_k)\|$ of the algorithm's output. In fact, we observe a perturbation behavior around $k = 63$, which, however, does not cause a breakdown of the algorithm. This again reflects the robustness of the used method.

**Example 3: Subspace tracking using the parameterized Newton algorithm**
Using the same setups as in the above numerical examples, we checked, if the parameterized Newton algorithm (Algorithm 2) is also able to track the time-varying principal subspace of $A(t)$. It turned out, that this is true for all considered examples, and in

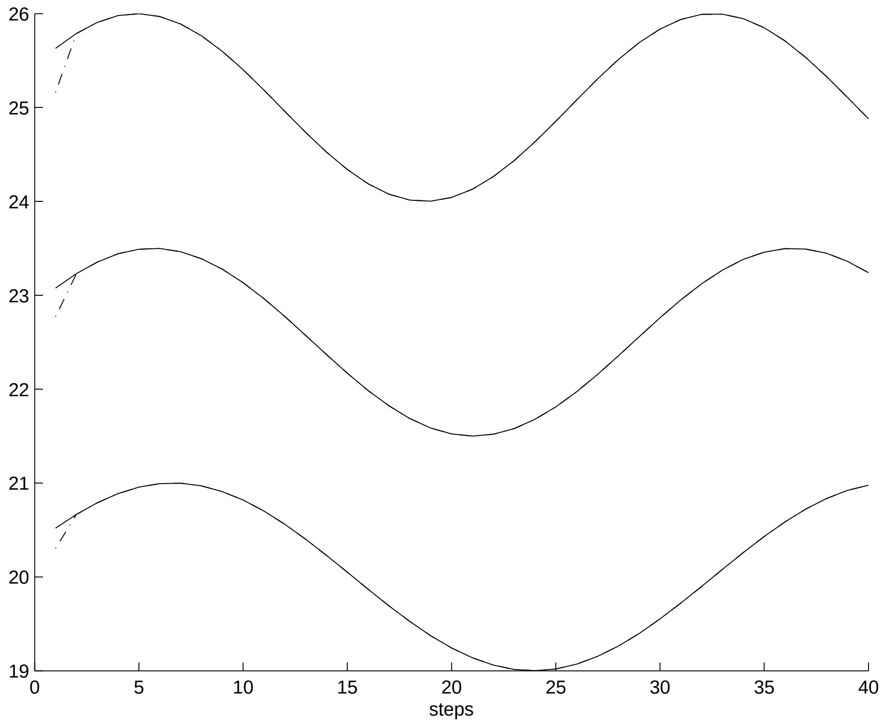


Figure 3.3: The eigenvalues of $[I_3\ 0]\Theta_k^\top A(t_k)\Theta_k[I_3\ 0]^\top$ (dotted) compared to the principal eigenvalues of $A(t_k)$, corresponding to Figure 3.1.

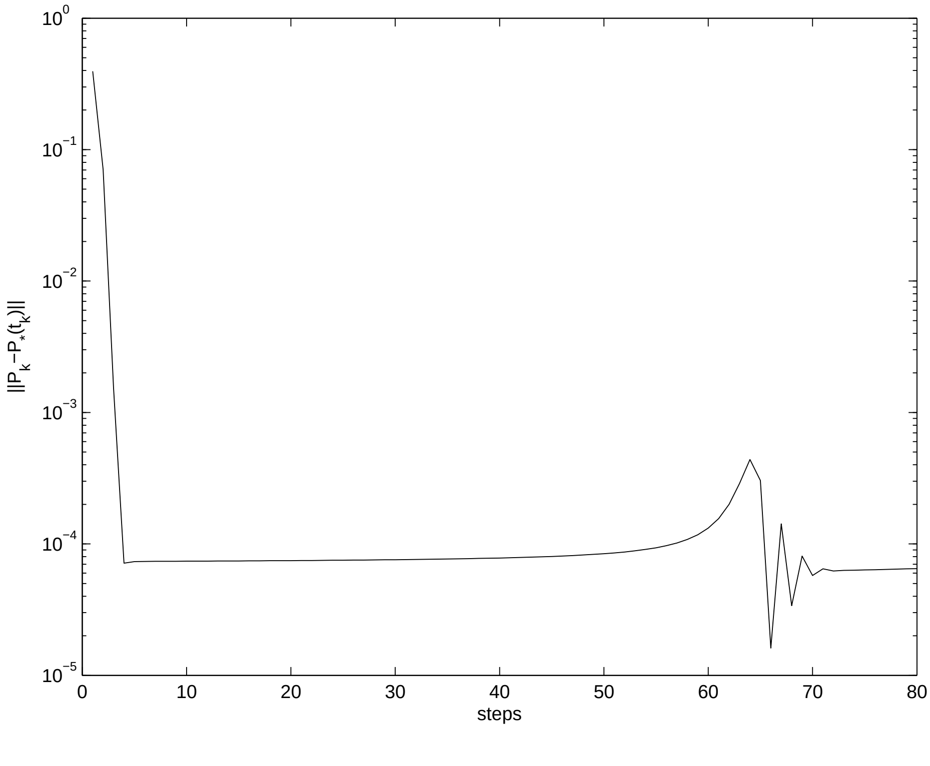


Figure 3.4: Almost coalescing eigenvalues: The error plot $\|P_k - P_*(t_k)\|$.

addition, the accuracy of the computed values was better than the results of Algorithm 4. E.g. for step size 0.025 and $A(t)$ as defined in Example 1, we observed an accuracy of $3.2 \cdot 10^{-6}$ which is significantly better than $1.2 \cdot 10^{-4}$ for Algorithm 4.
This is a phenomenon, which we have observed also in other simulations: if the situation is such, that the standard (parameterized) Newton method is also able to perform the tracking task, then the results have a better accuracy than those based on the time-varying approach. The main advantage of the time-varying Newton update scheme is, that it is more robust under perturbations and allows larger step sizes.
We therefore increased the step sizes to 0.1 and increased the time-dependency of the principal subspace by 5 times; i.e. we used $\tilde{X}_*(t) := X_*(5t)$ instead of $X_*(t)$ in equation (3.20). We computed 40 steps for 100 test runs using different random matrices $R$ in (3.21) and started with perfect initial conditions in each test.
Then we observed exactly what is expected: The time-varying Newton algorithm is still able to perform the tracking task with an accuracy of $1.1 \cdot 10^{-1}$ in all test runs. The parameterized Newton algorithm (Algorithm 2) showed in 75% of all cases a higher accuracy, but in the remaining 25%, the algorithm completely failed. The failure could be observed in the error plots, where the values arbitrarily varied between 0 and 1.
In order to combine the robustness of the time-varying Newton (Algorithm 4) with the accuracy of the parameterized Newton update scheme (Algorithm 2), we suggest to merge these methods. Thus Algorithm 4 is extended by an additional corrector step. The implementation is as follows:

**Algorithm 4': Extended parameterized time-varying Newton**

For $P_k \in \mathrm{Gr}_{m,n}$ and $\Theta_k \in O(n)$ with $P_k = \Theta_k \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta_k^\top$, compute $P_{k+1}$ and $\Theta_{k+1}$ in two steps:

1. Obtain estimates $P'_{k+1}$ and $\Theta'_{k+1}$ of $P_{k+1}$ and $\Theta_{k+1}$ by applying one step of Algorithm 4 (i.e. by evaluating equation (3.19)).

2. $P_{k+1}$ and $\Theta_{k+1}$ are given for $\begin{bmatrix} N_1 & N_2 \\ N_2^\top & N_3 \end{bmatrix} = (\Theta'_{k+1})^\top A(t_{k+1}) \Theta'_{k+1}$ by

$$P_{k+1} = \Theta_{k+1} \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta_{k+1}^\top \tag{3.22}$$

   and

$$\Theta_{k+1} = \Theta'_k \begin{bmatrix} I_m & -Z_k \\ Z_k^\top & I_{n-m} \end{bmatrix}_{\mathrm{Q}}$$

   where $Z_k \in \mathbb{R}^{m \times (n-m)}$ solves the Sylvester equation

$$N_1 Z_k - Z_k N_3 = N_2.$$

**Example 4: Comparison of different subspace tracking methods**

We have seen in the previous examples, that the methods based on the time-varying and time-invariant Riemannian Newton algorithm are well suitable for tracking. We

| | Riemannian Newton (Algorithm 1) | | Parameterized Newton (Algorithm 2) | | Extended parameterized time-varying Newton (Algorithm 4') | |
|---|---|---|---|---|---|---|
| n | Comp.time | Mean error | Comp.time | Mean error | Comp.time | Mean error |
| 10 | 0.06 | $5.7 \cdot 10^{-6}$ | 0.02 | $3.2 \cdot 10^{-6}$ | 0.06 | $2.4 \cdot 10^{-8}$ |
| 20 | 0.09 | $6.9 \cdot 10^{-6}$ | 0.06 | $1.2 \cdot 10^{-6}$ | 0.09 | $2.5 \cdot 10^{-8}$ |
| 40 | 0.28 | $5.6 \cdot 10^{-7}$ | 0.17 | $5.0 \cdot 10^{-7}$ | 0.38 | $2.9 \cdot 10^{-8}$ |
| 80 | 1.38 | $1.2 \cdot 10^{-6}$ | 0.92 | $1.3 \cdot 10^{-6}$ | 2.15 | $4.5 \cdot 10^{-7}$ |
| 160 | 8.99 | $6.3 \cdot 10^{-7}$ | 6.89 | $6.5 \cdot 10^{-7}$ | 14.6 | $6.2 \cdot 10^{-7}$ |

Table 3.2: The computing time and mean error of the algorithms for different $n$. We used $m = 3$.

therefore compare the tracking results of Algorithm 1, Algorithm 2 and Algorithm 4' with each other, where we observe the computing time and accuracy.
In order to do so, we used different $n, m \in \mathbb{N}$, step size $h = 0.025$, $t_k = kh$, $k = 1, ..., 40$, and

$$A(t) = X_*(t)K(t)X_*(t)^\top \in \mathbb{R}^{n\times n},$$

for $K(t) = \operatorname{diag}(a_n+\sin(nt), ..., a_1+\sin(1t))$, $X_*(t) = R^\top \begin{bmatrix} \cos(t) & \sin(t) & 0 \\ -\sin(t) & \cos(t) & 0 \\ 0 & 0 & I_{n-2} \end{bmatrix} R.$
Here, $R \in O(n)$ is a fixed random orthogonal matrix and $a_i := 2.5i$ for $i = 1, ..., n$.
As before, the task is to compute estimates $P_k$ of the principal subspace $P_*(t_k)$ of $A(t_k)$.
To measure the accuracy of the algorithm's output, we use the formula $\frac{1}{40}\sum_{k=1}^{40} \|P_k - P_*(t_k)\|$.
In Table 3.2, we see the computing time and accuracy of the algorithms for fixed $m = 3$ and perfect initial conditions, whereas in Table 3.3, we modified $m$ for fixed $n = 80$. We observe, that all algorithms are able to track the subspace of $A(t_k)$ at a reasonable accuracy. However, Algorithm 2 shows the best performance regarding the computing time and Algorithm 4' yields the most accurate results.

| | Riemannian Newton (Algorithm 1) | | Parameterized Newton (Algorithm 2) | | Extended parameterized time-varying Newton (Algorithm 4') | |
|---|---|---|---|---|---|---|
| m | Comp.time | Mean error | Comp.time | Mean error | Comp.time | Mean error |
| 3 | 1.38 | $1.2 \cdot 10^{-6}$ | 0.92 | $1.3 \cdot 10^{-6}$ | 2.15 | $4.5 \cdot 10^{-7}$ |
| 6 | 9.16 | $2.2 \cdot 10^{-6}$ | 3.70 | $1.9 \cdot 10^{-6}$ | 8.80 | $5.7 \cdot 10^{-7}$ |
| 12 | 67.2 | $1.4 \cdot 10^{-6}$ | 21.3 | $1.5 \cdot 10^{-6}$ | 39.4 | $2.1 \cdot 10^{-7}$ |
| 24 | 500 | $3.3 \cdot 10^{-6}$ | 82.4 | $4.2 \cdot 10^{-6}$ | 195 | $2.7 \cdot 10^{-7}$ |

Table 3.3: The computing time and mean error of the algorithms for different $m$. We used $n = 80$.

# Chapter 4

# Application II: Tracking Matrix Decompositions

In this chapter we apply the previously introduced tracking methods to the task of determining certain decompositions of time-varying matrices. At first we consider the eigenvalue decomposition (EVD) of time-varying symmetric matrices in the case of simple eigenvalues. Robust tracking algorithms for this purpose will be derived by using the time-varying Newton flow. Since it is a common situation to have symmetric matrices with distinct groups of eigenvalues of constant multiplicities, we will furthermore investigate this case. Then we are able to use a well known relation between the singular value decomposition (SVD) and the EVD to modify these algorithms such that they determine the time-varying SVD of non-square matrices. The resulting tracking algorithm then can be used to derive related matrix factorizations. Thus we give a SVD-based method to compute the polar decomposition of a family of invertible matrices. At the end of this chapter we reexamine the minor and principal eigenvector tracking of time-varying matrices.

In the following sections, some well known conditions guaranteeing the existence of a smooth SVD (EVD) of time-varying (symmetric) matrices will be cited. To get efficient tracking algorithms for the decomposition, however, the standard eigenvalue algorithms for constant matrices, see e.g. Horn [36] are not suited to track the desired orthogonal factors, since these methods are not designed to profit from the smoothness of the orthogonal factors.

Differential equations to track the time-varying EVD and SVD can be found in numerous publications, see e.g. Wright [70], Bell [9] or in the paper of Dieci [20], where the latter provides a good overview for such methods. By using discretization techniques, recursive update schemes can be derived from these ODEs. To guarantee a reasonable accuracy of the algorithm's output, however, one needs to execute intermediate corrector methods, since the discretization error accumulates at each step. Otherwise, the update scheme would produce senseless results after a number of steps, depending on the setup of the problem and the discretization technique chosen.

In contrast to this, the methods derived in this chapter for the EVD and SVD automatically perform error correction up to a certain accuracy. In particular, the tracking error does not accumulate at each step but remains at a fixed level, which is a major benefit of using an approach based on the time-varying Newton flow.

## 4.1 Eigenvalue and singular value decomposition

In this section we initially consider the problem of diagonalizing a smooth family of symmetric matrices $A(t)$. Thus, we want to determine a smooth map $t \mapsto X(t) \in O(n)$ such that for all $t \in \mathbb{R}$ holds

$$A(t) = X(t)D(t)X(t)^\top,$$

where $D(t)$ is a diagonal matrix. In the case, that $A(t)$ has distinct eigenvalues, the desired transformations are locally unique and can be determined via our theory by applying the standard (approximative) Newton flow to a suitable vector field, cf. [6]. Subsequently, we extend this first result to symmetric matrices with a group of equal eigenvalues. As the diagonalizing transformations are not uniquely determined then, we have to modify the original approach and use the tracking algorithm for underdetermined systems. Finally, these results are used to derive an update scheme, which performs the singular value decomposition of a sequence of non-square matrices.

### 4.1.1 Diagonalization of symmetric matrices with distinct eigenvalues

Let $t \mapsto A(t) \in \mathbb{R}^{n\times n}$, be a $C^r$-family ($r \geq 2$) of real symmetric matrices, with eigenvalues $\lambda_1(t), ..., \lambda_n(t)$, $t \in \mathbb{R}$. If these eigenvalues are distinct for all $t \in \mathbb{R}$, there exists a $C^r$-family of real orthogonal transformations $X_*(t)$, such that $X_*(t)^\top A(t)X_*(t)$ is diagonal for all $t \in \mathbb{R}$, cf. [62]. Our goal is to track such transformations by using the time-varying Newton flow.
In order to do so, we further require that $A(t)$ satisfies the following conditions:

1. $\|A(t)\|$, $\|\dot{A}(t)\|$ and $\|\ddot{A}(t)\|$ are uniformly bounded on $\mathbb{R}$.
2. There exists a constant $c > 0$ such that $|\lambda_i(t) - \lambda_j(t)| \geq c$ for $i \neq j$ and all $t \in \mathbb{R}$.

We now reformulate the original eigenvector tracking task into a zero-finding problem. Consider therefore the time-varying vector field

$$F : \mathbb{R}^{n\times n} \times \mathbb{R} \to \mathbb{R}^{n\times n} \tag{4.1}$$

defined by

$$F(X, t) = [N, X^\top A(t)X] + X^\top X - I, \tag{4.2}$$

where $I$ is the identity matrix, $N = \operatorname{diag}(1, ..., n)$ and $[,]$ is the Lie-Bracket product defined as $[A, B] := AB - BA$ for $A, B \in \mathbb{R}^{n\times n}$.

**Lemma 4.1.** *$F(X, t) = 0$ if and only if $X$ is an orthogonal matrix such that $X^\top A(t)X$ is diagonal.*

*Proof.* Note that the first summand of $F$ is skew symmetric while the second one is symmetric. Thus $F$ vanishes if and only if the two summands vanish, i.e. if and only if $X$ is orthogonal and

$$[N, X^\top A(t)X] = 0.$$

Since $N$ is diagonal with distinct eigenvalues, the result follows. □

Hence, the task of finding an orthogonal transformation $X$ such that $X^\top A(t)X$ is diagonal, is equivalent to that of finding a zero of $F(X, t)$. In order to use the time-varying Newton flow, certain technical assumptions have to be checked. This is done in the next lemma.

**Lemma 4.2.** *Let $A(t)$ and $F : \mathbb{R}^{n\times n} \times \mathbb{R} \to \mathbb{R}^{n\times n}$ be as above. There exists a continuously differentiable isolated solution $X_*(t)$ to $F(X, t) = 0$. $F(X, t)$ is $C^\infty$ in $X$ and $C^2$ in $(X, t)$. There exist constants $M_1, M_2, M_3, M_4 > 0$ such that*

1. $\|D^2F(X, t)\| \leq M_1,$ for all $X \in \mathbb{R}^{n\times n}$
2. $\|DF(X_*(t), t)\| \leq M_2,$
3. $\|\frac{\partial}{\partial t}DF(X, t)\| \leq M_3,$ for all $X \in B_r(X_*(t))$ *for some* $r > 0$.
4. $\|DF(X_*(t), t)^{-1}\| \leq M_4,$

*holds for all* $t \in \mathbb{R}$.

*Proof.* The claim concerning the differentiability properties of $F(X, t)$ is obvious. The first and second partial derivatives of $F$ w.r.to $X$ are the linear and bilinear maps $DF(X, t)$ and $D^2F(X, t)$ respectively, given as

$$DF(X, t) \cdot H = [N, H^\top A(t)X + X^\top A(t)H] + H^\top X + X^\top H, \tag{4.3}$$

$$D^2F(X, t) \cdot (H, \hat{H}) = [N, H^\top A(t)\hat{H} + \hat{H}^\top A(t)H] + H^\top \hat{H} + \hat{H}^\top H, \tag{4.4}$$

and the partial derivatives of $F(X, t)$ and $DF(X, t)$ w.r.to $t$ are

$$\frac{\partial}{\partial t}F(X, t) = [N, X^\top \dot{A}(t)X], \tag{4.5}$$

and

$$\frac{\partial}{\partial t}DF(X, t) \cdot H = [N, H^\top \dot{A}(t)X + X^\top \dot{A}(t)H]. \tag{4.6}$$

From this we deduce the operator norm estimates

$$\|DF(X, t)\| \leq 2(1 + 2a\|N\|)\|X\|, \tag{4.7}$$

$$\left\|\frac{\partial}{\partial t}DF(X, t)\right\| \leq 4d\|N\|\|X\|, \tag{4.8}$$

and $\|D^2F(X,t)\| \leq 2(1+2a\|N\|)$, where $a$ denotes the infinity-norm of $A$ and $d$ denotes the infinity-norm of $\dot{A}$. This shows 1)-3).
We next show, that the partial derivative operator $DF(X_*,t)$ is invertible for any solution $(X_*,t)$ of $F(X,t)=0$. In particular, $X_*$ is orthogonal and

$$DF(X_*,t)\cdot H = [N, H^\top A(t)X_* + X_*^\top A(t)H] + H^\top X_* + X_*^\top H. \tag{4.9}$$

Substituting $H = X_* \cdot \xi$, for $\xi \in \mathbb{R}^{n\times n}$ arbitrary, we obtain

$$\begin{aligned} DF(X_*,t)\cdot(X_*\xi) &= [N, \xi^\top X_*^\top A(t)X_* + X_*^\top A(t)X_*\xi] + \xi^\top X_*^\top X_* + X_*^\top X_*\xi. \quad (4.10)\\ &= [N, \xi^\top D + D\xi] + \xi^\top + \xi, \end{aligned}$$

where $D = X_*^\top A(t)X_*$ is diagonal. Thus $X_*\xi$ is in the kernel of $DF(X_*,t)$ if and only if $\xi$ is skew symmetric and $[N,[D,\xi]] = 0$. Hence $[D,\xi]$ must be diagonal and since $D$ has distinct diagonal entries we conclude that $\xi = 0$. This shows that $DF(X_*,t)$ is invertible for any zero of $F$. By the implicit function Theorem it follows that for every orthogonal $X_0$ with $X_0^\top A(0)X_0$ diagonal, there exists a unique $C^2$-curve $X_*(t)$ of orthogonal matrices with $X_*(0) = X_0$. This shows the first claim.
To prove 4), we derive a lower bound for the singular values of $DF(X_*,t)$. Let $\xi_{pq}$ denote the entry of $\xi$ with the largest absolute value. Assuming that the norm of $\xi$ is equal to one, the absolute value of $\xi_{pq}$ is at least $\frac{1}{n^2}$. The smallest singular value of $DF(X_*,t)$ is lower bounded by the sum of squares

$$(\xi_{pq}\lambda_p(p-q) + \xi_{qp}\lambda_q(p-q))^2 + (\xi_{pq}+\xi_{qp})^2$$

of the $pq$-entries of $[N, \xi^\top D + D\xi]$ and $\xi^\top + \xi$. For $p = q$ this is lower bounded by $\frac{4}{n^4}$, while otherwise it is lower bounded by $(\xi_{pq}\lambda_p + \xi_{qp}\lambda_q)^2 + (\xi_{pq}+\xi_{qp})^2$.
The latter is a quadratic function in $\xi_{qp}$ with minimum:

$$\frac{\xi_{pq}^2(\lambda_p-\lambda_q)^2}{1+\lambda_q^2} \geq \frac{(\lambda_q-\lambda_p)^2}{n^4(1+\lambda_q^2)}.$$

This is the desired lower bound for the singular values of $DF(X_*,t)$. Thus 4) follows with

$$M_4 = n^2 \max\left(\frac{1}{2}, \frac{1+a}{m}\right).$$

□

It follows, that $F$ satisfies the conditions of Theorem 2.9. Thus, the solution $X(t)$ of the differential equation

$$\dot{X} = DF(X,t)^{-1}\left(\mathcal{M}(X)F(X,t) - \frac{\partial}{\partial t}F(X,t)\right) \tag{4.11}$$

exists for all $t \in \mathbb{R}$ and converges exponentially to $X_*(t)$, provided $X(0)$ is sufficiently close to $X_*(0)$ and $\mathcal{M}$ is a stable bundle map.

Moreover, the necessary conditions of the discrete Newton flow are satisfied, cf. Theorem 2.10. Recall that the algorithms works at discrete times $t_k = kh$ for $h > 0$, $\mathcal{M}(X) = -\frac{1}{h}I$ and $k \in \mathbb{N}$.
Therefore, the sequence

$$X_{k+1} = X_k - DF(X_k, t_k)^{-1}\left(F(X_k, t_k) + hF_\tau^h(X_k, t_k)\right) \tag{4.12}$$

is well defined and produces estimates $X_k$ for $X_*(t_k)$, whose accuracy can be controlled by the step size. Here, the approximation $F_\tau^h(X, t)$ of $\frac{\partial}{\partial t}F(X, t)$ can be chosen as described in Section 2.1.2.2.

**Vectorizing the algorithm**
In order to implement the above algorithm, we need a way to compute the inverse of $DF$. One possibility is to employ the well known VEC operation and the Kronecker product, cf. appendix.
Consider the vectorization of equation (4.12)

$$\mathrm{VEC}(X_{k+1}) = \mathrm{VEC}(X_k) - H(X_k, t_k)^{-1}\mathrm{VEC}\left(F(X_k, t_k) + hF_\tau^h(X_k, t_k)\right), \tag{4.13}$$

where, $H$ is a matrix representation of $DF$, which can be determined by considering

$$H(X,t)\cdot\mathrm{VEC}(H) = \mathrm{VEC}\left([N, H^\top A(t)X + X^\top A(t)H] + H^\top X + X^\top H\right) =$$

$$\mathrm{VEC}\left(NH^\top A(t)X + NX^\top A(t)H - H^\top A(t)XN - X^\top A(t)HN + H^\top X + X^\top H\right) =$$

$$\big((X^\top A(t)\otimes N)\pi + I\otimes NX^\top A(t) - (NX^\top A(t)\otimes I)\pi - N\otimes X^\top A(t) +$$

$$+ (X^\top \otimes I)\pi + I\otimes X^\top\big)\,\mathrm{VEC}(H),$$

where $\otimes$ denotes the Kronecker product and $\pi \in \mathbb{R}^{n^2\times n^2}$ is such that for all $Z \in \mathbb{R}^{n\times n}$ holds

$$\pi\mathrm{VEC}(Z) = \mathrm{VEC}(Z^\top).$$

Hence, the matrix representation of $DF$ is given as

$$H(X,t) =$$

$$\left(X^\top A(t)\otimes N - NX^\top A(t)\otimes I + X^\top\otimes I\right)\pi + I\otimes NX^\top A(t) - N\otimes X^\top A(t) + I\otimes X^\top.$$

By noting the results from Lemma 4.2, the next theorem directly follows from Theorem 2.10.

**Theorem 4.1.** *Let $A(t)$, $N$ and $F : \mathbb{R}^{n\times n}\times\mathbb{R}\to\mathbb{R}^{n\times n}$ as above and let $X_*(t)$ denote a $C^2$-family of orthogonal matrices such that $F(X_*(t), t) = 0$ for all $t$. Let further $F_\tau^h(X, t)$ denote an approximation of $\frac{\partial}{\partial t}F(X, t)$ of order $p \geq 1$ and let*

$$H(X,t) :=$$

$$\left(X^\top A(t)\otimes N - NX^\top A(t)\otimes I + X^\top\otimes I\right)\pi + I\otimes NX^\top A(t) - N\otimes X^\top A(t) + I\otimes X^\top.$$

*Then the sequence $(X_k)_{k\in\mathbb{N}}$, defined by*

$$\boxed{\mathrm{VEC}(X_{k+1}) = \mathrm{VEC}(X_k) - H(X_k,t_k)^{-1}\mathrm{VEC}\left(F(X_k,t_k) + hF_\tau^h(X_k,t_k)\right).} \tag{4.14}$$

*satisfies for fixed $c > 0$ and sufficiently small $h > 0$*

$$\|X_k - X_*(t_k)\| \leq ch$$

*for all $k \in \mathbb{N}$, provided that $\|X_0 - X_*(0)\| < ch$.*

Note that the above defined algorithm needs the inversion of a matrix of size $n^2 \times n^2$ which implies, that this procedure is not practical for large $n$. Therefore, it will be studied in the following, how to approximatively invert the linear map $DF(X,t)$ without computing its matrix representation.

**Solving the Sylvester equation**
To replace the differential (4.11) equation by one of the form $\dot{X} = X\Omega$, consider

$$DF(X,t)\cdot X\Omega = [N, \Omega^\top X^\top A(t)X + X^\top A(t)X\Omega] + X^\top X\Omega + \Omega^\top X^\top X = K + Y, \tag{4.15}$$

where $K, Y$ are given skew symmetric and symmetric matrices, respectively.
Equation (4.15) is equivalent to the following equations

$$X^\top X\Omega + \Omega^\top X^\top X = Y, \tag{4.16}$$

$$[N, \Omega^\top X^\top A(t)X + X^\top A(t)X\Omega] = K. \tag{4.17}$$

According to [3], a general solution to (4.16) is given by

$$\Omega = (X^\top X)^{-1}\left(Z + \frac{1}{2}Y\right), \tag{4.18}$$

where $Z = -Z^\top$ is the only remaining variable. We insert this equation for $\Omega$ into (4.17) and obtain

$$\left[N, (-Z + \frac{1}{2}Y)X^{-1}A(t)X + X^\top A(t)(X^{-1})^\top(Z + \frac{1}{2}Y)\right] = K,$$

which is equivalent to

$$\begin{aligned} [N, -ZX^{-1}A(t)X + X^\top A(t)(X^{-1})^\top Z] &= \\ K - \frac{1}{2}[N, YX^{-1}A(t)X + X^\top A(t)(X^{-1})^\top Y]&. \end{aligned} \tag{4.19}$$

We use an approximation for $\frac{\partial}{\partial t}F(X,t)$, which is given by $F_\tau^h = [N, X^\top A_\tau^h(t)X]$, where $A_\tau^h(t)$ is a step size-dependent approximation of $\dot{A}(t)$. Note further, that in our situation, $K, Y$ are the skew symmetric and symmetric part of $-\frac{1}{h}F(X,t) - F_\tau^h(X,t)$, respectively. Thus, $K = [N, -\frac{1}{h}X^\top A(t)X - X^\top A^h_\tau(t)X]$, $Y = -\frac{1}{h}(X^\top X - I)$ and (4.19) can be written as

$$[N, -ZX^{-1}A(t)X + X^\top A(t)(X^{-1})^\top Z] = \tag{4.20}$$
$$[N, -\frac{1}{h}X^\top A(t)X - X^\top A^h_\tau(t)X - \frac{1}{2}(YX^{-1}A(t)X + X^\top A(t)(X^{-1})^\top Y)].$$

This yields the following condition for the off-diagonal entries:

$$-ZX^{-1}A(t)X + X^\top A(t)(X^{-1})^\top Z = R, \tag{4.21}$$

where $R = -\frac{1}{h}X^\top A(t)X - X^\top A^h_\tau(t)X - \frac{1}{2}(YX^{-1}A(t)X + X^\top A(t)(X^{-1})^\top Y)$.
Since $R$ is symmetric, this is a set of $n(n-1)/2$ linear equations in the entries of Z, which uniquely determines the skew symmetric matrix $Z$.
For $t \in \mathbb{R}$, let $X_* = X_*(t)$. Then (4.21) turns for $X = X_*$ into

$$-ZX_*^\top A(t)X_* + X_*^\top A(t)X_* Z = R_*, \tag{4.22}$$

$R_* = -\frac{1}{h}X_*^\top A(t)X_* - X_*^\top A^h_\tau(t)X_* - \frac{1}{2}(YX_*^\top A(t)X_* + X_*^\top A(t)X_* Y)$, and therefore

$$Z_{ij} = \begin{cases} \frac{(R_*)_{ij}}{\lambda_i - \lambda_j}, & i \neq j \\ 0, & i = j \end{cases} \tag{4.23}$$

This shows in particular, that the linear system (4.21) is solvable and well conditioned in $X = X_*$, since the eigenvalues $\lambda_i$ of $X_*^\top A(t)X_*$ are well separated. Thus (4.21) is robust under small changes of the left and right sides, for $X$ sufficiently close to $X_*$. For this reason, we replace in (4.21) $X^{-1}$ by the approximation $X^\top$ and $X^\top A(t)X$ by its diagonal part $D := \operatorname{diag}((X^\top A(t)X)_{11}, ..., (X^\top A(t)X)_{nn})$ and obtain an explicit formula for the approximation $\tilde{Z}$ of $Z$:

$$\tilde{Z}_{ij} = \begin{cases} \frac{R_{ij}}{D_{ii} - D_{jj}}, & i \neq j \\ 0, & i = j. \end{cases} \tag{4.24}$$

We arrive at the following tracking algorithm. Note that the corresponding algorithm of Dieci [20] only consists of the first summand $\Omega^e_{ij}$ of our algorithm. Thus, the additional terms stabilize the algorithm s.th. it is robust under perturbations.

**Theorem 4.2.** *Let $A(t)$ as above and let $A^h_\tau(t)$ denote an approximation of $\dot{A}(t)$ of order $p \geq 1$. Let further $t \mapsto X_*(t)$ denote a $C^2$ curve of orthogonal matrices such that $X_*(t)^\top A(t)X_*(t)$ is diagonal. Define for $X \in \mathbb{R}^{n\times n}$, $t \in \mathbb{R}$ and $h > 0$, $Y(X,t) := \frac{1}{h}(I - X^\top X)$, $d_i(X,t) := (X^\top A(t)X)_{ii}$ for $i = 1, ..., n$, $D(X,t) := \operatorname{diag}(d_1(X,t), ..., d_n(X,t))$ and*

$$\Omega(X,t) := \Omega^e(X,t) + \Omega^d(X,t) + \Omega^o(X,t),$$

*where*

1. $\Omega^e_{ij}(X,t) = \begin{cases} \frac{(X^\top A^h_\tau(t)X)_{ij}}{d_j(X,t) - d_i(X,t)} & , i \neq j, \\ 0 & , i = j. \end{cases}$

2. $\Omega^d_{ij}(X,t) = \begin{cases} \frac{(\frac{1}{h}X^\top A(t)X + \frac{1}{2}(Y(X,t)D(X,t)+D(X,t)Y(X,t)))_{ij}}{d_j(X,t)-d_i(X,t)} & , \ i \neq j \\ 0 & , \ i = j \end{cases}$

3. $\Omega^o(X,t) = \frac{1}{2}Y(X,t)$.

*Then for $c > 0$ and sufficiently small $h > 0$, the sequence*

$$\boxed{X_{k+1} = X_k + hX_k\Omega(X_k, t_k)} \tag{4.25}$$

*satisfies for $k \in \mathbb{N}$ and $t_k = kh$*

$$\|X_k - X_*(t_k)\| \leq ch,$$

*provided $X_0$ is sufficiently close to $X_*(0)$.*

*Proof.* We have derived an explicit expression $X(\tilde{Z} + \frac{1}{2}Y)$ for $G(X,t)\left(F^h_\tau(X,t) + \frac{1}{h}F(X,t)\right)$, where $F(X,t) = [N, X^\top A(t)X] + X^\top X - I$, $F^h_\tau(X,t) = [N, X^\top A^h_\tau(t)X]$ and $G$ denotes the used approximation for $DF^{-1}$.
In order to apply Theorem 2.4 for the derived tracking algorithm, we consider the perturbation term, which is given by

$$\Pi(X,t) = (I - DF(X,t)G(X,t))\left(F^h_\tau(X,t) + \frac{1}{h}F(X,t)\right). \tag{4.26}$$

We have to show, that

$$\|(\Pi(X,t))\| \leq c\|F(X,t)\|, \tag{4.27}$$

for some constant $c > 0$ and all $\|X - X_*(t)\| \leq r$, $t \geq 0$. By Lemma 4.2, there exist constants $M, R > 0$ such that for all $\|X - X_*(t)\| \leq R$, $t \geq 0$ holds

$$\|DF(X,t)\| \leq M. \tag{4.28}$$

Therefore, using (4.26)

$$\begin{aligned} \|\Pi(X,t)\| = \left\|DF(X,t)(DF(X,t)^{-1} - G(X,t))\left(F^h_\tau(X,t) + \frac{1}{h}F(X,t)\right)\right\| \leq \\ M\left\|(DF(X,t)^{-1} - G(X,t))\left(F^h_\tau(X,t) + \frac{1}{h}F(X,t)\right)\right\| = \\ M\left\|(X^\top)^{-1}(Z + \frac{1}{2}Y) - X(\tilde{Z} + \frac{1}{2}Y)\right\|, \end{aligned}$$

where $DF(X,t)^{-1}\left(-\frac{1}{h}F(X,t) - F^h_\tau(X,t)\right) = (X^\top)^{-1}(Z + \frac{1}{2}Y)$, cf. equation (4.18) and also $G(X,t)\left(-\frac{1}{h}F(X,t) - F^h_\tau(X,t)\right) = X(\tilde{Z} + \frac{1}{2}Y)$, since $G$ is an approximation of $DF^{-1}$.

We get that

$$\|\Pi(X,t)\| \leq M\left(\left\|\frac{1}{2}\left((X^\top)^{-1}-X\right)Y\right\| + \|X(Z-\tilde{Z}) + ((X^\top)^{-1}-X)Z\|\right), \quad (4.29)$$

where $Z$ and $\tilde{Z}$ are skew symmetric matrices satisfying

$$\text{offdiag }(-ZX^{-1}A(t)X + X^\top A(t)(X^\top)^{-1}Z) = \quad (4.30)$$

$$\text{offdiag }\left(-\frac{1}{h}X^\top A(t)X - X^\top A_\tau^h(t)X - \frac{1}{2}(YX^{-1}A(t)X + X^\top A(t)(X^{-1})^\top Y)\right)$$

and

$$\text{offdiag }(-\tilde{Z}\text{diag}(X^\top A(t)X) + \text{diag}(X^\top A(t)X)\tilde{Z}) = \quad (4.31)$$

$$\text{offdiag }\left(-\frac{1}{h}X^\top A(t)X - X^\top A_\tau^h(t)X - \frac{1}{2}(YX^\top A(t)X + X^\top A(t)XY)\right),$$

respectively. We assume that for some $c, h > 0$ and all $t \in \mathbb{R}$, $\|X - X_*(t)\| =: \delta \leq ch$. Then we can estimate the terms in (4.29) as follows:

1. $\|(X^\top)^{-1} - X\| \leq c_1\delta$ for some $c_1 > 0$:

   We use that $X = X_*(t)(I+E)$, where $\|E\| \leq \delta$. Thus

$$(X^\top)^{-1} - X = X_*(t)(I+E)^{-T}(I+E)^{-T} - X_*(t)(I+E) =$$

$$X_*(t)((I+E)^{-T} - I - E) = X_*(t)\left((I+E)^{-T} - (I+E)^{-T}(I+E)^T - E\right) =$$

$$X_*(t)\left((I+E)^{-T}(I-(I+E)^T) - E\right)$$

   Thus

$$\|(X^\top)^{-1} - X\| \leq \|(I+E)^{-T}E^T\| + \|E\| \leq \|E\|(\|(I+E)^{-1}\| + 1) \leq$$

$$\|E\|\left(\frac{1}{1-\|E\|} + 1\right) \leq \delta\left(\frac{1}{1-\delta} + 1\right),$$

   provided $\delta \leq 1$.

2. $\|Y\| \leq c_2$ for some $c_2 > 0$:

$$X^\top X = (I + E^\top)X_*(t)^\top X_*(t)(I+E) = I + E^\top E + E^\top + E$$

   Hence,

$$\|Y\| = \frac{1}{h}\|I - X^\top X\| = \frac{1}{h}\|E^\top E + E^\top + E\| \leq \frac{\delta}{h}(\delta+2) \leq c(\delta+2).$$

3. $\|Z\| \leq c_3$ for some $c_3 > 0$:
   $Z$ is the solution of the linear system (4.30), which is of the form
   $$L(X,t) \cdot Z = B(X,t).$$
   As mentioned before, this system is well conditioned in $X = X_*(t)$. Moreover, it is easily seen that $DL(X_*(t),t)$ and $D^2L(X,t)$ are uniformly bounded on $B_R(X_*(t))$, $t \in \mathbb{R}$. Thus $\|L(X,t)^{-1}\|$ is bounded for $X \in B_r(X_*(t))$ for some $r > 0$ and all $t \in \mathbb{R}$. Since $B(X,t) = \operatorname{offdiag}\left(-\frac{1}{h}X^\top A(t)X - X^\top A_\tau^h(t)X - \frac{1}{2}(YX^{-1}A(t)X + X^\top A(t)(X^{-1})^\top Y)\right)$ and the off-diagonal elements of $X^\top A(t)X$ and $hY$ are of order $h$, the norm of $B$ is bounded. This shows the boundedness of $\|Z\| = \|L(X,t)^{-1} \cdot B(X,t)\|$.

4. $\|Z - \tilde{Z}\| \leq c_4\delta$, for some $c_4 > 0$
   First notice that $\tilde{Z}$ is the solution of the linear system (4.31), which is of the form
   $$\tilde{L}(X,t) \cdot \tilde{Z} = \tilde{B}(X,t).$$
   Then we use the familiar estimate for solutions of linear systems (see e.g. [64]) and get:
   $$\|Z - \tilde{Z}\| \leq \|L^{-1}\|\|B - \tilde{B}\| + \frac{\|L^{-1}\|\|\tilde{B}\|\|L - \tilde{L}\|}{1 - \|L - \tilde{L}\|}.$$
   From this inequality we conclude the claim, as $\|B - \tilde{B}\| = O(\delta)$ and $\|L - \tilde{L}\| = O(\delta)$.

Thus (4.29) implies that
$$\|\Pi(X,t)\| \leq \delta M \left(\frac{1}{2}c_1c_2 + c_4\|X\| + c_1c_3\right)$$
for all $\|X - X_*(t)\| \leq ch$, $t \geq 0$. Thus
$$\|\Pi(X,t)\| \leq \delta M \left(\frac{1}{2}c_1c_2 + c_4\frac{1}{1-ch} + c_1c_3\right).$$
Note that under the assumptions on $F$, we have that $\|F(X,t)\| \geq \kappa\|X - X_*(t)\|$ for some $\kappa, r > 0$ for all $t$ and $X \in B_r(X_*(t))$ (cf. Theorem 2.12, claim 1).
We get that
$$\|\Pi(X,t)\| \leq \frac{M}{\kappa}\|F(X,t)\| \left(\frac{1}{2}c_1c_2 + c_4\frac{1}{1-ch} + c_1c_3\right).$$
This shows that the necessary conditions of Theorem 2.12 are satisfied, which proves the stability claim for (4.25). □

Note that $X_*(t)$ is no longer uniquely defined, if $A(t)$ has multiple eigenvalues, as eigenspaces of dimension $> 1$ occur. Thus, the tracking algorithm of Theorem 4.2 is not applicable in this form and needs further modifications. These are derived in the following section.

### 4.1.2 Diagonalization of symmetric matrices with multiple eigenvalues

An important extension of the previously introduced EVD tracking algorithms for time-varying symmetric matrices is to derive methods, working also with multiple eigenvalues. This is an obvious demand, since it is a natural situation to have one or more groups of identical eigenvalues. We therefore make the following assumptions for the time-varying matrix $A(t)$.

**A1** The map $t \mapsto A(t) \in \text{Sym}(n)$ is $C^r$ with $r \geq 2$.

**A2** For some fixed $m \in \mathbb{N}, m < n$, the eigenvalues of $A(t)$ satisfy for all $t \in \mathbb{R}$

$$\lambda_1(t) \neq ... \neq \lambda_{m+1}(t)$$

and

$$\lambda_{m+1}(t) = ... = \lambda_n(t).$$

**A3** The norms $\|A(t)\|$, $\|\dot{A}(t)\|$ and $\|\ddot{A}(t)\|$ are uniformly bounded on $\mathbb{R}$.

**A4** There exists a constant $c > 0$ such that for all $t \in \mathbb{R}$ holds

$$|\lambda_i(t) - \lambda_j(t)| \geq c, \quad \text{for } i \neq j \text{ and } 1 \leq i, j \leq m+1.$$

The next proposition shows, that under the conditions A1 and A2 , there exists a $C^r$-map $t \mapsto Q(t) \in O(n)$, such that $Q(t)^\top A(t) Q(t) = \text{diag}(\lambda_1(t), ..., \lambda_n(t))$ for all $t \in \mathbb{R}$.

**Proposition 4.1.** *Let $A(t) \in \mathbb{R}^{n \times n}$, $t \in \mathbb{R}$ denote a family of symmetic matrices with eigenvalues $\lambda_1(t), ..., \lambda_n(t)$.*

1. *If $t \mapsto A(t)$ is $C^r$, $r \geq 1$, and $\lambda_i(t) \neq \lambda_j(t)$ for $i \neq j$ and $t \in \mathbb{R}$, then there exists a $C^r$-eigenvalue decomposition, i.e. there exists a $C^r$-family of orthogonal matrices $Q(t)$, such that $Q(t)^\top A(t) Q(t)$ is diagonal.*

2. *In the case that $A(t)$ has $q$ groups of identical eigenvalues, we subsumize the identical eigenvalues by defining $\lambda_1(t) = ... = \lambda_{p_1}(t) =: \Lambda_1(t)$,..., $\lambda_{p_{q-1}+1}(t) = ... = \lambda_{p_{p_q}}(t) =: \Lambda_q(t)$ for $1 \leq p_1 < ... < p_q = n$ for some $1 \leq q \leq n$.*

   *If $t \mapsto A(t)$ is $C^r$, $r \geq 1$, and $\Lambda_i(t) \neq \Lambda_j(t)$ for $i \neq j$ and $t \in \mathbb{R}$, then there exists a $C^r$-eigenvalue decomposition of $A(t)$.*

3. *If $A(t)$ is real analytic in $t$, then there exist a real analytic eigenvalue decomposition.*

*Proof.* The first and second claim can be found in Sibuya [62], while the last one has been shown by Kato [41]. $\square$

Our goal is to track a diagonalizing $C^r$-transformation $Q(t)$ of $A(t)$, which, however, is neither unique nor isolated: Any curve

$$\tilde{Q}(t) := Q(t) \begin{bmatrix} I_m & \\ & R \end{bmatrix},$$

where $R \in O(n-m)$ yields an other diagonalizing transformation of $A(t)$. But the first $m$ columns of $Q(t)$ are unique except for their sign. It is therefore appropriate to define the set of curves, which contain $Q(t)$ and whose elements all diagonalize $A(t)$:

$$\mathcal{X}_*(t) := \left\{ Q(t) \begin{pmatrix} I_m & \\ & R \end{pmatrix} \Bigg| \; R \in O(n-m) \right\}. \tag{4.32}$$

Thus, $X^\top A(t) X$ is diagonal for all $X \in \mathcal{X}_*(t)$, $t \in \mathbb{R}$. Other orthogonal matrices diagonalizing $A(t)$ are well separated from $\mathcal{X}_*(t)$, since they have at least one different sign in one of the first $m$ columns.
In the sequel, we construct a dynamical system, whose solution $X(t)$ converges exponentially to $\mathcal{X}_*(t)$, i.e.

$$\operatorname{dist}(X(t), \mathcal{X}_*(t)) \leq ae^{-bt} \quad \text{for some } a, b > 0.$$

We therefore consider the map

$$F : \mathbb{R}^{n\times m} \times \mathbb{R}^{n\times(n-m)} \times \mathbb{R} \to \begin{pmatrix} \mathbb{R}^{m\times m} & \mathbb{R}^{m\times(n-m)} \\ \mathbb{R}^{(n-m)\times m} & \mathrm{Sym}(n-m) \end{pmatrix},$$

defined by

$$F(X_1, X_2, t) = \left[ N, \begin{pmatrix} X_1^\top A(t) X_1 & X_1^\top A(t) X_2 \\ X_2^\top A(t) X_1 & 0 \end{pmatrix} \right] + (X_1 \; X_2)^\top (X_1 \; X_2) - I_n, \tag{4.33}$$

where $I_n$ is the identity matrix, $N = \operatorname{diag}(1, ..., n)$, $[,]$ denotes the Lie bracket and $(X_1 \; X_2)$ denotes the $n \times n$ matrix, which results from concatenating $X_1$ and $X_2$.
The next lemma shows that the zeros of $F(\cdot, t)$ are the desired diagonalizing transformations of $A(t)$.

**Lemma 4.3.** *Let $A(t)$ satisfy the assumptions A1-A4, let $F(X_1, X_2, t)$ and $\mathcal{X}_*(t)$ as above and let $(X_1^* \; X_2^*) \in \mathcal{X}_*(t)$ for some $t \in \mathbb{R}$. Then for sufficiently small $r > 0$ and $X_1 \in B_r(X_1^*)$ holds that $F(X_1, X_2, t) = 0$ if and only if $(X_1 \; X_2)$ is orthogonal and $X_1 = X_1^*$.*

*Proof.* As the first summand of $F$ is skew symmetric and the second is symmetric, $F$ vanishes only, if the summands themselves vanish. The second one is zero, if and only if $(X_1, X_2)$ is orthogonal, while the first one vanishes, if $N$ commutes with $G := \begin{pmatrix} X_1^\top A(t) X_1 & X_1^\top A(t) X_2 \\ X_2^\top A(t) X_1 & 0 \end{pmatrix}$, i.e. $G$ is diagonal. This is fulfilled, if the columns of $(X_1 \; X_2)$ are pairwise orthogonal eigenvectors of $A(t)$. In particular, the columns of $X_1$ are the eigenvectors of $A(t)$ to the pairwise distinct eigenvalues $\lambda_1(t), ..., \lambda_m(t)$, whose multiplicity is 1. They are therefore locally unique, i.e. $X_1 = X_1^*$. Since the eigenspace of $\lambda_{m+1}(t)$ has dimension $n - m$, $X_2$ can be any matrix such that $(X_1 \; X_2)$ is orthogonal. $\square$

In order to track the zero of $F$, we want to employ the time-varying Newton flow and need therefore the following result.

**Lemma 4.4.** *Let $A(t)$ satisfy the assumptions A1-A4, let $F(X_1, X_2, t)$ and $\mathcal{X}_*(t)$ as above. Then there exist some $M_1, M_2, M_3, M_4 > 0$ such that following statements hold:*

1. $\|D^2F(X_1, X_2, t)\| \leq M_1$, *for all* $(X_1, X_2) \in \mathbb{R}^{n\times m} \times \mathbb{R}^{n\times(n-m)}$, $t \in \mathbb{R}$.
2. $\|DF(X_1^*, X_2^*, t)\| \leq M_2$, *for all* $(X_1^* \ X_2^*) \in \mathcal{X}_*(t)$, $t \in \mathbb{R}$.
3. $\|\frac{\partial}{\partial t} DF(X_1, X_2, t)\| \leq M_3$, *for all* $(X_1 \ X_2) \in U_r(t) = \{X \in \mathbb{R}^{n\times n} | \operatorname{dist}(X, \mathcal{X}_*(t)) < r\}$, $t \in \mathbb{R}$ *and some fixed* $r > 0$.
4. $\ker(DF(X_1^*, X_2^*, t)) = \left\{ (X_1^* \ X_2^*) \begin{pmatrix} 0 & 0 \\ 0 & Z \end{pmatrix} \middle| Z \in \mathfrak{so}(\mathrm{n}-\mathrm{m}) \right\}$, *for all* $(X_1^* \ X_2^*) \in \mathcal{X}_*(t)$, $t \in \mathbb{R}$.
5. $\mathrm{rk} DF(X_1^*, X_2^*, t) = n^2 - \frac{1}{2}(n-m)(n-m-1)$, *for all* $(X_1^* \ X_2^*) \in \mathcal{X}_*(t)$, $t \in \mathbb{R}$.
6. $\sigma_{min}(DF(X_1^*, X_2^*, t)^\top) \geq M_4$ *for all* $(X_1^* \ X_2^*) \in \mathcal{X}_*(t)$, $t \in \mathbb{R}$.

*Here, $DF$ and $D^2F$ denote the derivatives of $F$ with respect to $(X_1 \ X_2)$ and $\sigma_{min}(B)$ denotes the smallest singular value of $B \in \mathbb{R}^{n\times m}$.*

*Proof.* Consider at first the derivatives of $F$ with respect to $(X_1 \ X_2) =: X$, where $H := (H_1 \ H_2)$ denotes a tangent vector of $\mathbb{R}^{n\times m} \times \mathbb{R}^{n\times(n-m)}$.

$$DF(X_1, X_2, t) \cdot H = DF(X_1, X_2, t) \cdot (H_1 \ H_2) =$$

$$\left[ N, \begin{pmatrix} H_1^\top A(t) X_1 + X_1^\top A(t) H_1 & H_1^\top A(t) X_2 + X_1^\top A(t) H_2 \\ H_2^\top A(t) X_1 + X_2^\top A(t) H_1 & 0 \end{pmatrix} \right] + H^\top X + X^\top H.$$

$$D^2F(X_1, X_2, t) \cdot ((H_1 \ H_2), (\hat{H}_1 \ \hat{H}_2)) =$$

$$\left[ N, \begin{pmatrix} H_1^\top A(t) \hat{H}_1 + \hat{H}_1^\top A(t) H_1 & H_1^\top A(t) \hat{H}_2 + \hat{H}_1^\top A(t) H_2 \\ H_2^\top A(t) \hat{H}_1 + \hat{H}_2^\top A(t) H_1 & 0 \end{pmatrix} \right] + H^\top \hat{H} + \hat{H}^\top H.$$

Moreover,

$$\frac{\partial}{\partial t} DF(X_1, X_2, t) \cdot (H_1 \ H_2) = \left[ N, \begin{pmatrix} H_1^\top \dot{A} X_1 + X_1^\top \dot{A} H_1 & H_1^\top \dot{A} X_2 + X_1^\top \dot{A} H_2 \\ H_2^\top \dot{A} X_1 + X_2^\top \dot{A} H_1 & 0 \end{pmatrix} \right].$$

Thus, the claims 1)-3) are obvious, as $\|(X_1 \ X_2)\|$, $\|A(t)\|$ and $\|\dot{A}(t)\|$ are uniformly bounded for $(X_1 \ X_2) \in U_r(t)$, $t \in \mathbb{R}$.
To show 4) we use for tangent vectors $H =: X\xi$ with $X = (X_1 \ X_2)$ and $\xi = (\xi_1 \ \xi_2)$ and consider

$$DF(X_1, X_2, t) \cdot H = DF(X_1, X_2, t) \cdot (H_1 \ H_2) =$$

$$\left[ N, \begin{pmatrix} H_1^\top A(t) X_1 + X_1^\top A(t) H_1 & H_1^\top A(t) X_2 + X_1^\top A(t) H_2 \\ H_2^\top A(t) X_1 + X_2^\top A(t) H_1 & 0 \end{pmatrix} \right] + H^\top X + X^\top H =$$

$$\left[N, \begin{pmatrix} \xi_1^\top X^\top A(t)X_1 + X_1^\top A(t)X\xi_1 & \xi_1^\top X^\top A(t)X_2 + X_1^\top A(t)X\xi_2 \\ \xi_2^\top X^\top A(t)X_1 + X_2^\top A(t)X\xi_1 & 0 \end{pmatrix}\right] + \xi^\top X^\top X + X^\top X\xi.$$

Let $D = \operatorname{diag}(\lambda_1(t), ..., \lambda_n(t))$ and consider for $t \in \mathbb{R}$ the derivative of $F$ in $X = (X_1^* \ X_2^*) \in \mathcal{X}_*(t)$:

$$DF(X_1^*, X_2^*, t) \cdot (X_1^* \ X_2^*)(\xi_1 \ \xi_2) =$$

$$\left[N, \begin{pmatrix} \xi_1^\top (X_*)^\top A(t)X_1^* + (X_1^*)^\top A(t)X_*\xi_1 & \xi_1^\top (X_*)^\top A(t)X_2^* + (X_1^*)^\top A(t)X_*\xi_2 \\ \xi_2^\top (X_*)^\top A(t)X_1^* + (X_2^*)^\top A(t)X_*\xi_1 & 0 \end{pmatrix}\right] + \xi^\top + \xi$$

$$=: K + Y$$

This is a sum of a skew symmetric ($K$) and a symmetric term ($Y$), which is zero, if each summand vanishes. To have $Y = 0$ we need $\xi^\top = -\xi$. Now consider $K$ and use $D_1 = \operatorname{diag}(\lambda_1(t), ..., \lambda_m(t))$, $\hat{D}_1 := \begin{pmatrix} D_1 \\ 0 \end{pmatrix} \in \mathbb{R}^{n\times m}$ and $D_2 = \operatorname{diag}(\lambda_{m+1}(t), ..., \lambda_{m+1}(t))$ , $\hat{D}_2 := \begin{pmatrix} 0 \\ D_2 \end{pmatrix} \in \mathbb{R}^{n\times(n-m)}$:

$$K = \left[N, \begin{pmatrix} \xi_1^\top \hat{D}_1 + \hat{D}_1^\top \xi_1 & \xi_1^\top \hat{D}_2 + \hat{D}_1^\top \xi_2 \\ \xi_2^\top \hat{D}_1 + \hat{D}_2^\top \xi_1 & 0 \end{pmatrix}\right].$$

By using $\xi_1 = \begin{pmatrix} \xi_{11} \\ \xi_{21} \end{pmatrix}$ and $\xi_2 = \begin{pmatrix} \xi_{12} \\ \xi_{22} \end{pmatrix}$, where $\xi_{11} \in \mathbb{R}^{m\times m}$ and $\xi_{22} \in \mathbb{R}^{(n-m)\times(n-m)}$ are skew symmetric and $\xi_{12}^\top = -\xi_{21} \in \mathbb{R}^{(n-m)\times m}$, the above equation is equivalent to

$$K = \left[N, \begin{pmatrix} \xi_{11}^\top D_1 + D_1\xi_{11} & \xi_{21}^\top D_2 + D_1\xi_{12} \\ \xi_{12}^\top D_1 + D_2\xi_{21} & 0 \end{pmatrix}\right].$$

Recall that $N$ only commutes with diagonal matrices. Hence, it is a necessary condition that the skew symmetric (sub-)matrix $\xi_{11} = 0$ to have $K = 0$, as the diagonal entries of $D_1$ are distinct. Now consider the position $(i, j)$ of the matrix $\xi_{21}^\top D_2 + D_1\xi_{12}$:

$$(\xi_{21}^\top D_2 + D_1\xi_{12})_{ij} = -(\xi_{12})_{ij}(D_2)_{jj} + (D_1)_{ii}(\xi_{12})_{ij} =$$

$$-(\xi_{12})_{ij}\lambda_{m+1} + \lambda_i(\xi_{12})_{ij} = (\xi_{12})_{ij}(\lambda_i - \lambda_{m+1}).$$

This shows, that $K = 0$ only if $\xi_{12} = 0$. Hence, the only degree of freedom is the choice of $\xi_{22} = -\xi_{22}^\top$ and the kernel of $DF$ is given by

$$\ker(DF(X_1^*, X_2^*, t)) = \left\{ (X_1^* \ X_2^*) \begin{pmatrix} 0 & 0 \\ 0 & Z \end{pmatrix} \middle| \ Z \in \mathfrak{so}(\mathrm{n-m}) \right\}.$$

Thus,

$$\dim\ker(DF(X_1^*, X_2^*, t)) = \frac{1}{2}(n-m)(n-m-1),$$

which shows 4).

5) follows immediately from 4).

6) To show the last claim, let $H \perp \ker(DF(X_1^*, X_2^*, t))$ and $\|H\| = 1$. $H$ can be written as

$$H = X_* \xi =: X_* \begin{pmatrix} \xi_{11} & \xi_{12} \\ \xi_{21} & \xi_{22} \end{pmatrix}.$$

As $H$ is orthogonal to the kernel of $DF$, we must have $\xi_{22} = \xi_{22}^\top$, cf. above. Let $(i,j)$ denote the position of the biggest entry of $\xi$. As $\|H\| = 1$ and $X_*$ is orthogonal we have $\|\xi\| = 1$ and thus $|\xi_{ij}| \geq \frac{1}{n^2}$. With the notation of above, we have

$$DF(X_1^*, X_2^*, t) \cdot X_* \xi = K + \xi + \xi^\top,$$

where $K^\top = -K$ and

$$K = \left[ N, \begin{pmatrix} \xi_{11}^\top D_1 + D_1 \xi_{11} & \xi_{21}^\top D_2 + D_1 \xi_{12} \\ \xi_{12}^\top D_1 + D_2 \xi_{21} & 0 \end{pmatrix} \right].$$

Hence,

$$K = \left[ N, \begin{pmatrix} D_1 & \\ & D_2 \end{pmatrix} \begin{pmatrix} \xi_{11} & \xi_{12} \\ \xi_{21} & 0 \end{pmatrix} + \begin{pmatrix} \xi_{11}^\top & \xi_{21}^\top \\ \xi_{12}^\top & 0 \end{pmatrix} \begin{pmatrix} D_1 & \\ & D_2 \end{pmatrix} \right].$$

Note that $\xi + \xi^\top$ is symmetric and therefore

$$\|DF(X_1^*, X_2^*, t) \cdot X_* \xi\|^2 = \|K\|^2 + \|\xi + \xi^\top\|^2,$$

since $K \perp (\xi^\top + \xi)$. To lower bound the smallest singular value of $DF(X_1^*, X_2^*, t)^\top$, we consider the entry of $DF(X_1^*, X_2^*, t) \cdot X_* \xi$ at position $(i,j)$ and distinguish two cases:

(i) $i \leq m \vee j \leq m$

Then

$$K_{ij} = (i - j)(\lambda_i \xi_{ij} + \lambda_j \xi_{ji}),$$

where $\lambda_i$ denotes the $i$th eigenvalue of $A(t)$. Therefore,

$$(DF(X_1^*, X_2^*, t) \cdot X_* \xi)_{ij} = (i - j)(\lambda_i \xi_{ij} + \lambda_j \xi_{ji}) + \xi_{ij} + \xi_{ji}.$$

Let $s$ denote the smallest singular value of $DF^\top$. Thus

$$s^2 \geq (i - j)^2 (\lambda_i \xi_{ij} + \lambda_j \xi_{ji})^2 + (\xi_{ij} + \xi_{ji})^2.$$

If $i \neq j$, then

$$s^2 \geq (\lambda_i \xi_{ij} + \lambda_j \xi_{ji})^2 + (\xi_{ij} + \xi_{ji})^2,$$

which is a quadratic function in $\xi_{ji}$ for $|\xi_{ji}| \leq |\xi_{ij}|$ with minimum

$$\frac{\xi_{ij}^2 (\lambda_i - \lambda_j)^2}{1 + \lambda_j^2} \geq \frac{(\lambda_j - \lambda_i)^2}{n^4 (1 + \lambda_j^2)}.$$

For $i = j$, then $s^2 \geq (\xi_{ii} + \xi_{ii})^2 \geq \frac{4}{n^4}$.

(ii) $i > m \wedge j > m$

Here $K_{ij} = 0$ and since the (22) block of $\xi$ $(= \xi_{22})$ is symmetric we have that

$$(DF(X_1^*, X_2^*, t) \cdot X_*\xi)_{ij} = \xi_{ij} + \xi_{ji} = 2\xi_{ij} \geq \frac{2}{n^2}.$$

Thus the smallest singular value $s$ of $DF^\top$ satisfies $s^2 \geq \frac{4}{n^4}$.

$\square$

The previous lemma shows that $F$ satisfies the conditions of Theorem 2.13. Thus with $X = (X_1\ X_2)$, the solution $X(t) = (X_1(t)\ X_2(t))$ of the differential equation

$$\dot{X} = DF(X_1, X_2, t)^\dagger \left( \mathcal{M}(X) F(X_1, X_2, t) - \frac{\partial}{\partial t} F(X_1, X_2, t) \right) \tag{4.34}$$

exists for all $t \in \mathbb{R}$ and converges exponentially to $\mathcal{X}_*(t)$, provided $X(0)$ is sufficiently close to $\mathcal{X}_*(0)$. Here $DF^\dagger$ denotes a pseudo inverse of $DF$, cf. Chapter 2.3.3.

Furthermore, the discrete algorithm of Theorem 2.4 is applicable, which works at discrete times $t_k = kh$ for $h > 0$ and $k \in \mathbb{N}$. Hence for $X_k = (X_k^1\ X_k^2)$, the sequence

$$X_{k+1} = X_k - DF(X_1^k, X_2^k, t_k)^\dagger \left( F(X_1^k, X_2^k, t_k) + h F_\tau^h(X_1^k, X_2^k, t_k) \right) \tag{4.35}$$

is well defined and the accuracy of $X_k$ can be controlled by the step size. Note that the approximation $F_\tau^h(X_1, X_2, t)$ of $\frac{\partial}{\partial t} F(X_1, X_2, t)$ is given by

$$F_\tau^h(X_1, X_2, t) = \left[ N, \begin{pmatrix} X_1^\top A_\tau^h(t) X_1 & X_1^\top A_\tau^h(t) X_2 \\ X_2^\top A_\tau^h(t) X_1 & 0 \end{pmatrix} \right],$$

where $A_\tau^h(t)$ is an approximation of $\dot{A}(t)$.
Analogously to the previous section, one now has to vectorize the matrices $X_k, X_{k+1}$ and $(F(X_1^k, X_2^k, t_k) + hF_\tau^h(X_1^k, X_2^k, t_k))$ and compute the pseudo-inverse of a matrix representation of $DF$, which can be done by employing the VEC-operation and the Kronecker product. However, the matrix associated with $DF$ has the size $\left(n^2 - \frac{(n-m)(n-m-1)}{2}\right) \times \left(n^2 - \frac{(n-m)(n-m-1)}{2}\right)$, which shows, that this way of implementing the algorithm is not really practical. Thus, we now concentrate on solving the implicit linear equation, associated with (4.35), without computing a matrix representation of $DF$.

**Approximatively solving the implicit equation.**
The implicit form of (4.34) can be written for $\mathcal{M}(X) = -\frac{1}{h} I$ as

$$DF(X_1, X_2, t) \cdot \dot{X} = -\frac{1}{h} F(X_1, X_2, t) - F_\tau^h(X_1, X_2, t). \tag{4.36}$$

Let $K$ and $Y$ denote the skew symmetric and symmetric part of $-\frac{1}{h} F(X_1, X_2, t) - F_\tau^h(X_1, X_2, t)$, respectively. Then the above equation reads

$$DF(X_1, X_2, t) \cdot \dot{X} = K + Y. \tag{4.37}$$

We now determine an approximative solution of equation (4.37) and then formulate the tracking algorithm by using Theorem 2.4.
Let for $t \in \mathbb{R}$, $X_* = (X_1^* \ X_2^*)$ such that

$$X_*^\top A(t) X_* = \operatorname{diag}(\lambda_1(t), ..., \lambda_n(t)).$$

Then the derivative of $F$ with respect to $X := (X_1 \ X_2)$ is a linear map, which acts on elements of the tangent space $(X_1 \ X_2)(\xi_1 \ \xi_2)$ in the following manner

$$DF(X_1, X_2, t) \cdot (X_1 \ X_2)(\xi_1 \ \xi_2) =$$
$$\left[N, \begin{pmatrix} \xi_1^\top X^\top A(t) X_1 + X_1^\top A(t) X \xi_1 & \xi_1^\top X^\top A(t) X_2 + X_1^\top A(t) X \xi_2 \\ \xi_2^\top X^\top A(t) X_1 + X_2^\top A(t) X \xi_1 & 0 \end{pmatrix}\right] + \xi^\top X^\top X + X^\top X \xi$$

Thus for $\dot{X} = (X_1 \ X_2)(\xi_1 \ \xi_2)$ equation (4.37) can be written as

$$\left[N, \begin{pmatrix} \xi_1^\top X^\top A(t) X_1 + X_1^\top A(t) X \xi_1 & \xi_1^\top X^\top A(t) X_2 + X_1^\top A(t) X \xi_2 \\ \xi_2^\top X^\top A(t) X_1 + X_2^\top A(t) X \xi_1 & 0 \end{pmatrix}\right] + \xi^\top X^\top X + X^\top X \xi$$
$$= K + Y.$$

Inspecting the symmetry properties of the occurring terms shows, that the above equation is satisfied, if

$$\left[N, \begin{pmatrix} \xi_1^\top X^\top A(t) X_1 + X_1^\top A(t) X \xi_1 & \xi_1^\top X^\top A(t) X_2 + X_1^\top A(t) X \xi_2 \\ \xi_2^\top X^\top A(t) X_1 + X_2^\top A(t) X \xi_1 & 0 \end{pmatrix}\right] = K, \tag{4.38}$$

and

$$\xi^\top X^\top X + X^\top X \xi = Y. \tag{4.39}$$

A solution to (4.39) is given as

$$\xi = (X^\top X)^{-1} \left(Z + \frac{1}{2} Y\right), \tag{4.40}$$

for an arbitrary skew symmetric matrix $Z$. To get a solution of (4.38), consider this equation for $X = X_* = [X_1^* \ X_2^*]$ and obtain

$$\left[N, \begin{pmatrix} \xi_1^\top \hat{D}_1 + \hat{D}_1^\top \xi_1 & \xi_1^\top \hat{D}_2 + \hat{D}_1^\top \xi_2 \\ \xi_2^\top \hat{D}_1 + \hat{D}_2^\top \xi_1 & 0 \end{pmatrix}\right] = K, \tag{4.41}$$

where $\hat{D}_1 = X_*^\top A(t) X_1^* = \begin{pmatrix} D_1 \\ 0 \end{pmatrix}$, $D_1 = (X_1^*)^\top A(t) X_1^*$ and $\hat{D}_2 = X_*^\top A(t) X_2^* = \begin{pmatrix} 0 \\ D_2 \end{pmatrix}$, $D_2 = (X_2^*)^\top A(t) X_2^*$. By using $\xi_1 = \begin{pmatrix} \xi_{11} \\ \xi_{21} \end{pmatrix}$ and $\xi_2 = \begin{pmatrix} \xi_{12} \\ \xi_{22} \end{pmatrix}$, equation (4.38) simplifies to

$$\left[N, \begin{pmatrix} \xi_{11}^\top D_1 + D_1 \xi_{11} & \xi_{21}^\top D_2 + D_1 \xi_{12} \\ \xi_{12}^\top D_1 + D_2 \xi_{21} & 0 \end{pmatrix}\right] = K. \tag{4.42}$$

Due to equation (4.40), $\xi = Z + \frac{1}{2}Y$ for $Z^\top = -Z$ in $X = X_*$. Thus

$$\begin{pmatrix} \xi_{11} & \xi_{12} \\ \xi_{21} & \xi_{22} \end{pmatrix} = \begin{pmatrix} Z_{11} & Z_{12} \\ -Z_{12}^\top & Z_{22} \end{pmatrix} + \frac{1}{2}\begin{pmatrix} Y_{11} & Y_{12} \\ Y_{12}^\top & Y_{22} \end{pmatrix}.$$

Plug this into (4.42) and get

$$\begin{aligned} &\left[N, \begin{pmatrix} -Z_{11}D_1 + D_1Z_{11} & -Z_{12}D_2 + D_1Z_{12} \\ Z_{12}^\top D_1 - D_2Z_{12}^\top & 0 \end{pmatrix}\right] = \\ &K - \frac{1}{2}\left[N, \begin{pmatrix} Y_{11}D_1 + D_1Y_{11} & Y_{12}D_2 + D_1Y_{12} \\ Y_{12}^\top D_1 + D_2Y_{12}^\top & 0 \end{pmatrix}\right]. \end{aligned} \tag{4.43}$$

As $K$ is the skew symmetric part of $-\frac{1}{h}F(X_1, X_2, t) - F_\tau^h(X_1, X_2, t)$, we can use that

$$K = -\left[N, \frac{1}{h}\begin{pmatrix} X_1^\top A(t)X_1 & X_1^\top A(t)X_2 \\ X_2^\top A(t)X_1 & 0 \end{pmatrix} + \begin{pmatrix} X_1^\top A_\tau^h(t)X_1 & X_1^\top A_\tau^h(t)X_2 \\ X_2^\top A_\tau^h(t)X_1 & 0 \end{pmatrix}\right],$$

where $A_\tau^h$ denotes an approximation of $\dot{A}$. Therefore, (4.43) reads

$$\begin{aligned} &\left[N, \begin{pmatrix} -Z_{11}D_1 + D_1Z_{11} & -Z_{12}D_2 + D_1Z_{12} \\ Z_{12}^\top D_1 - D_2Z_{12}^\top & 0 \end{pmatrix}\right] = \\ &-\left[N, \begin{pmatrix} X_1^\top(\frac{1}{h}A(t) + A_\tau^h(t))X_1 & X_1^\top(\frac{1}{h}A(t) + A_\tau^h(t))X_2 \\ X_2^\top(\frac{1}{h}A(t) + A_\tau^h(t))X_1 & 0 \end{pmatrix} + \right. \\ &\left. \frac{1}{2}\begin{pmatrix} Y_{11}D_1 + D_1Y_{11} & Y_{12}D_2 + D_1Y_{12} \\ Y_{12}^\top D_1 + D_2Y_{12}^\top & 0 \end{pmatrix}\right] \end{aligned} \tag{4.44}$$

If we assume, $(X_*\xi) \perp \ker(DF(X_1^*, X_2^*, t))$, then $Z_{22} = 0$, cf. Lemma 4.4. Thus, the above equation uniquely determines the non-zero entries of the skew symmetric matrix $Z$:

$$Z_{ij} = \begin{cases} \frac{R_{ij}}{D_{ii} - D_{jj}}, & \text{for } i \neq j \text{ and } i \leq m \vee j \leq m \\ 0, & \text{else.} \end{cases} \tag{4.45}$$

Here, $D = X_*^\top A(t)X_* = \operatorname{diag}(\lambda_1(t), ..., \lambda_n(t))$ and

$$\begin{aligned} R = &-\begin{pmatrix} X_1^\top(\frac{1}{h}A(t) + A_\tau^h(t))X_1 & X_1^\top(\frac{1}{h}A(t) + A_\tau^h(t))X_2 \\ X_2^\top(\frac{1}{h}A(t) + A_\tau^h(t))X_1 & 0 \end{pmatrix} - \\ &\frac{1}{2}\begin{pmatrix} Y_{11}D_1 + D_1Y_{11} & Y_{12}D_2 + D_1Y_{12} \\ Y_{12}^\top D_1 + D_2Y_{12}^\top & 0 \end{pmatrix}. \end{aligned}$$

This shows that the solution $\dot{X} = X\xi$ of (4.37) is uniquely determined in $X = X_*$, if one assumes that $\dot{X}$ is orthogonal to $\ker(DF(X_1^*, X_2^*, t))$. Moreover, this system is well conditioned in $X = X_*$, as the differences $|\lambda_i(t) - \lambda_j(t)|$ are assumed to be lower bounded for $i \neq j$ and $i \leq m \vee j \leq m$.
Therefore, the solutions of (4.38) and (4.39) are robust under relatively small changes of the entries. To be able to solve these equations for $X \neq X_*$, we replace $X^\top X$ by $I$ in

(4.40) and in (4.38) we approximate $X^\top A(t)X_1$ by $\begin{pmatrix} \tilde{D}_1 \\ 0 \end{pmatrix}$, $\tilde{D}_1 = \operatorname{diag}((X^\top A(t)X_1)_{1,1},$ $..., (X^\top A(t)X_1)_{m,m})$ and $X^\top A(t)X_2$ by $\begin{pmatrix} 0 \\ \tilde{D}_2 \end{pmatrix}$ where $\tilde{D}_2 = \operatorname{diag}((X^\top A(t)X_2)_{m+1,1}, ...,$ $(X^\top A(t)X_1)_{n,(n-m)})$.

We therefore approximate $\dot{X} = (X^\top)^{-1}\left(Z + \frac{1}{2}Y\right)$ by $X\left(\tilde{Z} + \frac{1}{2}Y\right)$, for $X$ sufficiently close to $\mathcal{X}_*(t)$, where $\tilde{Z} = -\tilde{Z}^\top$ is given for $D = \operatorname{diag}((X^\top A(t)X)_{11}, ..., (X^\top A(t)X)_{nn})$ by

$$\tilde{Z}_{ij} = \begin{cases} \frac{R_{ij}}{D_{ii}-D_{jj}}, & \text{for } i \neq j \text{ and } i \leq m \vee j \leq m \\ 0, & \text{else.} \end{cases} \tag{4.46}$$

Note that we can determine the relevant entries of $R$ by the equation

$$R_{ij} = \hat{R}_{ij} \quad \text{for } i \leq m \vee j \leq m,$$

where $\hat{R} = -X^\top(\frac{1}{h}A(t) + A_\tau^h(t))X - \frac{1}{2}(YD + DY)$. The use of this formula simplifies the explicit expression of the update scheme in the following theorem.

Note that also in the case of multiple eigenvalues, a similar algorithm can be found in [20], which however lacks the corrector terms $\Omega_{ij}^d$, $\Omega_{ij}^0$ in the formula.

**Theorem 4.3.** *Let $A(t)$ satisfy the assumptions A1-A4, let $\mathcal{X}_*(t)$ as defined in* (4.32) *and let $A_\tau^h(t)$ denote an approximation of $\dot{A}(t)$ of order $p \geq 1$. Define for $X \in \mathbb{R}^{n\times n}$, $t \in \mathbb{R}$ and $h > 0$, $Y(X,t) := \frac{1}{h}(I - X^\top X)$, $d_i(X,t) := (X^\top A(t)X)_{ii}$, $D(X,t) :=$ $\operatorname{diag}(d_1(X,t), ..., d_n(X,t))$ and*

$$\Omega(X,t) := \Omega^e(X,t) + \Omega^d(X,t) + \Omega^o(X,t),$$

*where*

1. $\Omega_{ij}^e(X,t) = \begin{cases} \frac{(X^\top A_\tau^h(t)X)_{ij}}{d_j(X,t)-d_i(X,t)} & , i \neq j \text{ and } i \leq m \vee j \leq m \\ 0 & , \text{else.} \end{cases}$

2. $\Omega_{ij}^d(X,t) = \begin{cases} \frac{(\frac{1}{h}X^\top A(t)X + \frac{1}{2}(Y(X,t)D(X,t)+D(X,t)Y(X,t)))_{ij}}{d_j(X,t)-d_i(X,t)} & , i \neq j,\ i \leq m \vee j \leq m \\ 0 & , \text{else} \end{cases}$

3. $\Omega^o(X,t) = \frac{1}{2}Y(X,t)$.

*Then for $c > 0$ and sufficiently small $h > 0$, the sequence*

$$\boxed{X_{k+1} = X_k + hX_k\Omega(X_k, t_k)} \tag{4.47}$$

*satisfies for $k \in \mathbb{N}$ and $t_k = kh$*

$$\operatorname{dist}(X_k, \mathcal{X}_*(t_k)) \leq ch,$$

*provided $X_0$ is sufficiently close to $\mathcal{X}_*(0)$.*

*Proof.* We have derived an explicit expression $X(\tilde{Z}+\frac{1}{2}Y)$ for

$$G(X_1,X_2,t)\left(F_\tau^h(X_1,X_2,t)+\frac{1}{h}F(X_1,X_2,t)\right),$$

where $X=(X_1\ X_2)$,

$$F(X_1,X_2,t)=\left[N,\begin{pmatrix} X_1^\top A(t)X_1 & X_1^\top A(t)X_2 \\ X_2^\top A(t)X_1 & 0 \end{pmatrix}\right]+(X_1\ X_2)^\top(X_1\ X_2)-I_n,$$

$$F_\tau^h(X_1,X_2,t)=\left[N,\begin{pmatrix} X_1^\top A_\tau^h(t)X_1 & X_1^\top A_\tau^h(t)X_2 \\ X_2^\top A_\tau^h(t)X_1 & 0 \end{pmatrix}\right]$$

and $G$ denotes the used approximation for $DF^\dagger$.
In order to apply Theorem 2.4 for the derived tracking algorithm, we consider the perturbation term, which is given by

$$\Pi(X_1,X_2,t)= \tag{4.48}$$
$$(I-DF(X_1,X_2,t)G(X_1,X_2,t))\left(\frac{1}{h}F(X_1,X_2,t)+F_\tau^h(X_1,X_2,t)\right).$$

We have to show, that

$$\|(\Pi(X_1,X_2,t))\|\leq c\|F(X_1,X_2,t)\|, \tag{4.49}$$

for some constant $c>0$ and all $\|X-X_*(t)\|\leq r$, $t\geq 0$.
By Lemma 4.4, there exist constants $M,R>0$ such that for all $\|X-X_*(t)\|\leq R$, $t\geq 0$

$$\|DF(X_1,X_2,t)\|\leq M. \tag{4.50}$$

Therefore, using (4.48)

$$\begin{aligned}\|\Pi(X_1,X_2,t)\| &= \Big\|DF(X_1,X_2,t)(DF(X_1,X_2,t)^\dagger-G(X_1,X_2,t))\cdot \\ &\left(F_\tau^h(X_1,X_2,t)+\frac{1}{h}F(X_1,X_2,t)\right)\Big\|\leq \\ M&\left\|(DF(X_1,X_2,t)^\dagger-G(X_1,X_2,t))\left(F_\tau^h(X_1,X_2,t)+\frac{1}{h}F(X_1,X_2,t)\right)\right\|= \\ M&\left\|(X^\top)^{-1}(Z+\frac{1}{2}Y)-X(\tilde{Z}+\frac{1}{2}Y)\right\|,\end{aligned}$$

where, cf. equation (4.40), $DF(X_1,X_2,t)^\dagger\left(-\frac{1}{h}F(X_1,X_2,t)-F_\tau^h(X_1,X_2,t)\right)=(X^\top)^{-1}$ $(Z+\frac{1}{2}Y)$ and $G(X_1,X_2,t)\cdot\left(-\frac{1}{h}F(X_1,X_2,t)-F_\tau^h(X_1,X_2,t)\right)=X(\tilde{Z}+\frac{1}{2}Y)$.
Thus, we have the same situation as in the proof of Theorem 4.2 and the rest proceeds as there.

□

### 4.1.3 Singular value decomposition

We now consider the task of tracking the singular value decomposition (SVD) of a $C^r$-family ($r \geq 2$) of real matrices $t \mapsto M(t) \in \mathbb{R}^{m\times n}$, where $m \geq n$.
Thus, we want to determine a family of left singular vectors $V_*(t) \in O(m)$ and right singular vectors $U_*(t) \in O(n)$ of $M(t)$ such that

$$V_*^\top(t)M(t)U_*(t) = \hat{\Sigma}(t),$$

where

$$\hat{\Sigma}(t) = \begin{bmatrix} \Sigma(t) \\ 0_{(m-n)\times n} \end{bmatrix}, \quad \Sigma(t) = \mathrm{diag}\left(\sigma_1(t), ..., \sigma_n(t)\right).$$

Note that in the case $m > n$, the orthogonal factor $V_*(t)$ is not unique. To see this, we use the partition $V_*(t) =: [V_1^*(t)\ V_2^*(t)]$, where $V_1^* \in \mathbb{R}^{m\times n}$, $V_2^*(t) \in \mathbb{R}^{m\times(m-n)}$ and consider

$$[V_1^*(t)\ V_2^*(t)]^\top M(t)U_*(t) = \begin{bmatrix} V_1^*(t)^\top M(t)U_*(t) \\ V_2^*(t)^\top M(t)U_*(t) \end{bmatrix} = \begin{bmatrix} \Sigma(t) \\ 0_{(m-n)\times n} \end{bmatrix}.$$

Hence, for any $R \in O(m-n)$, the matrix

$$\tilde{V}_*(t) := V_*(t) \begin{bmatrix} I & \\ & R \end{bmatrix} = [V_1^*(t)\ V_2^*(t)R]$$

is an other orthogonal matrix such that

$$\tilde{V}_*^\top(t)M(t)U_*(t) = \hat{\Sigma}(t).$$

Therefore, we define for given $V_*(t) = [V_1^*(t)\ V_2^*(t)]$ a family of sets of orthogonal matrices

$$\mathcal{V}_*(t) := \left\{ [V_1^*(t)\ V_2^*(t)R] \middle|\ R \in O(m-n) \right\},$$

whose elements are all left singular vectors of $M(t)$.

It is a well known fact, that the singular value decomposition of $M(t)$ can be obtained by determining an orthogonal diagonalizing transformation $\hat{V}_*(t)$ of the matrix

$$\hat{M}(t) := \begin{bmatrix} 0 & M(t) \\ M(t)^\top & 0 \end{bmatrix},$$

such that

$$\hat{V}(t)_*^\top \hat{M}(t)\hat{V}_*(t) = \begin{bmatrix} \Sigma(t) & & \\ & -\Sigma(t) & \\ & & 0 \end{bmatrix}.$$

Then the orthogonal factors $U_*(t)$, $V_*(t) = [V_1^*(t)\ V_2^*(t)]$ for the SVD of $M(t)$ are given by

$$\hat{V}_*(t) = \frac{1}{\sqrt{2}} \begin{bmatrix} V_1^*(t) & V_1^*(t) & \sqrt{2}V_2^*(t) \\ U_*(t) & -U_*(t) & 0 \end{bmatrix} \in \mathbb{R}^{(m+n)\times(m+n)},$$

where $V_1^*(t) \in \mathbb{R}^{m\times n}$ and $V_2^*(t) \in \mathbb{R}^{m\times(m-n)}$. In the case that the non-zero eigenvalues of $\hat{M}(t)$ are pairwise distinct, the columns of $V_1^*(t)$ and $U_*(t)$ are uniquely determined, up to their sign. Thus the (locally unique) set of orthogonal matrices, which diagonalize $\hat{M}(t)$, and which contains $\hat{V}_*(t)$, is given by

$$\mathcal{W}_*(t) = \left\{ \frac{1}{\sqrt{2}} \left[ \begin{array}{ccc} V_1^*(t) & V_1^*(t) & \sqrt{2}V_2^*(t)R \\ U_*(t) & -U_*(t) & 0 \end{array} \right] \middle|\ R \in O(m-n) \right\}. \tag{4.51}$$

We now derive the SVD algorithm by using the EVD-results of the previous chapter for $\hat{M}(t)$. In order to guarantee the applicability of Theorem 4.3, we have to assume, that $\hat{M}(t)$ satisfies the assumptions A1-A4 made in the previous chapter, which implies, that we have to impose the following conditions on $M(t)$:

**M1** The map $t \mapsto M(t) \in \mathbb{R}^{m\times n}$, $m \geq n$, is $C^r$ with $r \geq 2$.

**M2** The singular values $\sigma_1(t), ..., \sigma_n(t)$ of $M(t)$ satisfy $\sigma_i(t) \neq \sigma_j(t)$ for $i \neq j$ and $t \in \mathbb{R}$.

**M3** $\|M(t)\|$, $\|\dot{M}(t)\|$ and $\|\ddot{M}(t)\|$ are uniformly bounded on $\mathbb{R}$.

**M4** There exists $c > 0$ such that the singular values of $M(t)$ satisfy for all $t \in \mathbb{R}$

$$|\sigma_i(t) - \sigma_j(t)| \geq c, \quad i \neq j.$$

The following result shows that a $C^r$-singular value decomposition of $M(t)$ already exists under weaker assumptions than M1 and M2.

**Proposition 4.2.** *Let $M(t) \in \mathbb{R}^{m\times n}$, $m \geq n$ and let $\sigma_1(t), ..., \sigma_n(t)$ denote the singular values of $M(t)$, $t \in \mathbb{R}$. In the case that $M(t)$ has $q$ groups of identical singular values, we define $\sigma_1(t) = ... = \sigma_{p_1}(t) =: \hat{\sigma}_1(t)$,..., $\sigma_{p_{q-1}+1}(t) = ... = \sigma_{p_q}(t) =: \hat{\sigma}_q(t)$ for $1 \leq p_1 < ... < p_q = n$ and some $q \in \mathbb{N}$.*

1. *If $t \mapsto M(t)$ is $C^r$, $r \geq 1$, and $\sigma_i(t) \neq \sigma_j(t)$ for $i \neq j$ and $t \in \mathbb{R}$, then there exists a $C^r$-singular value decomposition, i.e. there exists $C^r$-families of orthogonal matrices $V(t) \in \mathbb{R}^{m\times m}$, $U(t) \in \mathbb{R}^{n\times n}$ such that $V(t)^\top M(t)U(t) = \hat{\Sigma}(t)$ where $\hat{\Sigma}(t) = \left[ \begin{array}{c} \mathrm{diag}\,(\sigma_1(t), ..., \sigma_n(t)) \\ 0_{n-m} \end{array} \right]$.*

2. *If $t \mapsto M(t)$ is $C^r$, $r \geq 1$, and has $q$ groups of identically singular values satisfying $\hat{\sigma}_i(t) \neq \hat{\sigma}_j(t)$ for $i \neq j$ and all $t \in \mathbb{R}$, then there exists a $C^r$-singular value decomposition of $M(t)$.*

3. *If $M(t)$ is real analytic, then there exist a real analytic singular value decomposition.*

*Proof.* Claim 1 and 2 can be found in a publication of Dieci [20], while the 3rd has been shown by Bunse-Gerstner [14]. □

Note that in the case of dealing with the real analytic singular value decomposition, we would have to allow negative singular values. Thus, we would have to modify assumption M2 and M4 such that the absolute values of the singular values are assumed to be different.
However, we assume M1-M4 and formulate the SVD tracking theorem for a $C^r$ family of matrices for $r \geq 2$. This result is a generalization of the algorithm given in [6], since the new method also works for non-square matrices $M(t)$.

**Theorem 4.4.** *Let $M(t)$ satisfy M1-M4, let $U_*(t)$ and $\mathcal{V}_*(t)$ as above and let $M_\tau^h(t)$ denote an approximation of $\frac{\partial}{\partial t}M(t)$ of order $p \geq 1$. Let further for $V_1 \in \mathbb{R}^{m\times n}$, $V_2 \in \mathbb{R}^{m\times(m-n)}$, $U \in \mathbb{R}^{n\times n}$, $t \in \mathbb{R}$ and $h > 0$,*

$$Y_1(V_1, U) = \frac{1}{2}(2I_n - V_1^\top V_1 - U^\top U),$$

$$Y_2(V_1, U) = \frac{1}{2}(U^\top U - V_1^\top V_1),$$

$$Y_3(V_1, V_2) = -\frac{\sqrt{2}}{2}V_1^\top V_2,$$

$$Y_4(V_2) = I_{m-n} - V_2^\top V_2,$$

$$B_1(V_1, U, t) =$$
$$\frac{1}{2}\left(U^\top(hM_\tau^h(t) + M(t))^\top V_1 + V_1^\top(hM_\tau^h(t) + M(t))U + D(t)Y_1(V_1, U) + Y_1(V_1, U)D(t)\right),$$
$$B_2(V_1, U, t) =$$
$$\frac{1}{2}\left(U^\top(hM_\tau^h(t) + M(t))^\top V_1 - V_1^\top(hM_\tau^h(t) + M(t))U + D(t)Y_2(V_1, U) - Y_2(V_1, U)D(t)\right),$$

$$B_3(V_1, V_2, U, t) = \frac{\sqrt{2}}{2}U^\top\left(hM_\tau^h(t) + M(t)\right)^\top V_2 + \frac{1}{2}D(V_1, U, t)Y_3(V_1, V_2),$$

*where*

$$D(V_1, U, t) := \operatorname{diag}\left((U^\top M(t)^\top V_1)_{11}, ..., (U^\top M(t)^\top V_1)_{nn}\right),$$

*and $d_i(V_1, U, t) = (D(V_1, U, t))_{ii}$, $i = 1, ..., n$. Let moreover $Z_1(V_1, U, t) = -Z_1(V_1, U, t)^\top \in \mathbb{R}^{n\times n}$, $Z_2(V_1, U, t) = -Z_2(V_1, U, t)^\top \in \mathbb{R}^{n\times n}$ and $Z_3(V_1, V_2, U, t) \in \mathbb{R}^{n\times(m-n)}$ be defined as*

$$(Z_1(V_1, U, t))_{ij} = \begin{cases} (B_1(V_1, U, t))_{ij}(d_j(V_1, U, t) - d_i(V_1, U, t))^{-1}, & i \neq j \\ 0, & \text{else} \end{cases}$$

$$(Z_2(V_1, U, t))_{ij} = -(B_2(V_1, U, t))_{ij}(d_j(V_1, U, t) + d_i(V_1, U, t))^{-1}, \quad i, j = 1, ..., n$$

$$(Z_3(V_1, V_2, U, t))_{ij} = -(B_3(V_1, V_2, U, t))_{ij}d_i(V_1, U, t)^{-1}, \quad i = 1, ..., n, j = 1, ..., m-n$$

*Then for $c > 0$ and sufficiently small $h > 0$, the sequence $(V_k, U_k)$, defined for $V_k := [V_1^k\ V_2^k]$ by*

$$\boxed{\begin{aligned} V_1^{k+1} &= V_1^k + V_1^k \left( Z_1(V_1^k, U_k, t_k) + Z_2(V_1^k, U_k, t_k) + \tfrac{1}{2}(Y_1(V_1^k, U_k) + Y_2(V_1^k, U_k)) \right) + \\ &\quad \sqrt{2} V_2^k \left( -Z_3(V_1^k, V_2^k, U_k, t_k)^\top + \tfrac{1}{2} Y_3(V_1^k, V_2^k)^\top \right), \\ V_2^{k+1} &= V_2^k + \tfrac{1}{2} \left( \sqrt{2} V_1^k (2 Z_3(V_1^k, V_2^k, U_k, t_k) + Y_3(V_1^k, V_2^k)) + V_2^k Y_4(V_2^k) \right), \\ U_{k+1} &= U_k + U_k \left( Z_1(V_1^k, U_k, t_k) + \tfrac{1}{2} Y_1(V_1^k, U_k) - Z_2(V_1^k, U_k, t_k) - \tfrac{1}{2} Y_2(V_1^k, U_k) \right), \end{aligned}}$$

*satisfies for all* $k \in \mathbb{N}$

$$\operatorname{dist}(V_k, \mathcal{V}_*(t_k))^2 + \|U_k - U_*(t_k)\|^2 \leq c^2 h^2,$$

*provided* $(V_0, U_0)$ *is sufficiently close to* $(\mathcal{V}_*(0), U_*(0))$.

*Proof.* To use Theorem 4.3, let $\hat{V}_k = \frac{1}{\sqrt{2}} \begin{bmatrix} V_1^k & V_1^k & \sqrt{2} V_2^k \\ U_k & -U_k & 0 \end{bmatrix}$, $Y = \frac{1}{h}(I - \hat{V}_k^\top \hat{V}_k)$, $\hat{M}(t) = \begin{bmatrix} 0 & M(t) \\ M(t)^\top & 0 \end{bmatrix}$, $\hat{D} := \operatorname{diag}\left((\hat{V}_k \hat{M}(t_k) \hat{V}_k)_{1,1}, ..., (\hat{V}_k \hat{M}(t_k) \hat{V}_k)_{2n,2n}, 0, ..., 0\right) \in \mathbb{R}^{(n+m)\times(m+n)}$ and $\hat{d}_i = \hat{D}_{ii}$ for $i = 1, ..., n$

Let $\hat{M}_\tau^h(t) = \begin{bmatrix} 0 & M_\tau^h(t) \\ M_\tau^h(t)^\top & 0 \end{bmatrix}$ such that $B := \hat{V}_k^\top (\hat{M}_\tau^h(t_k) + \frac{1}{h} \hat{M}(t_k)) \hat{V}_k + \frac{1}{2}(Y \hat{D} + \hat{D} Y)$, let $\mathcal{W}_*(t)$ as defined in (4.51) and let $Z \in \mathbb{R}^{(m+n)\times(m+n)}$ be defined by

$$Z_{ij} := \begin{cases} B_{ij}(\hat{d}_j - \hat{d}_i)^{-1}, & i \neq j \text{ and } i \leq 2n \wedge j \leq 2n \\ 0, & \text{else.} \end{cases}$$

Then for $c > 0$ and sufficiently small $h > 0$, the sequence

$$\hat{V}_{k+1} = \hat{V}_k + h \hat{V}_k \left( Z + \frac{1}{2} Y \right) \tag{4.52}$$

satisfies for $k \in \mathbb{N}$

$$\operatorname{dist}(\hat{V}_k, \mathcal{W}_*(t_k)) \leq ch,$$

if $\hat{V}_0$ is sufficiently close to $\mathcal{W}_*(0)$, cf. Theorem 4.3.

Due to the fact, that each $\hat{V}_k$ consists of the orthogonal factors for the SVD of $M(t_k)$, we derive the claimed update scheme from the finding above. We therefore study the structure of the terms in (4.52):

Let therefore $\hat{V} = \frac{1}{\sqrt{2}} \begin{bmatrix} V_1 & V_1 & \sqrt{2} V_2 \\ U & -U & 0 \end{bmatrix}$ and we get that

$$\hat{V}^\top \hat{M} \hat{V} = \frac{1}{2} \begin{bmatrix} U^\top M^\top V_1 + V_1^\top M U & U^\top M^\top V_1 - V_1^\top M U & \sqrt{2} U^\top M^\top V_2 \\ V_1^\top M U - U^\top M^\top V_1 & -U^\top M^\top V_1 - V_1^\top M U & -\sqrt{2} U^\top M^\top V_2 \\ \sqrt{2} V_2^\top M U & -\sqrt{2} V_2^\top M U & 0 \end{bmatrix}.$$

Thus for $D := \operatorname{diag}((U^\top M^\top V_1)_{11}, ..., (U^\top M^\top V_1)_{nn})$, the diagonal entries of $\hat{V}^\top \hat{M} \hat{V}$ are given by

$$\hat{D} := \begin{bmatrix} D & & \\ & -D & \\ & & 0 \end{bmatrix}.$$

We consider

$$hY = I - \hat{V}^\top \hat{V} =$$

$$I - \frac{1}{\sqrt{2}} \begin{bmatrix} V_1^\top & U^\top \\ V_1^\top & -U^\top \\ \sqrt{2}V_2^\top & 0 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} V_1 & V_1 & \sqrt{2}V_2 \\ U & -U & 0 \end{bmatrix} =$$

$$\frac{1}{2} \begin{bmatrix} 2I - V_1^\top V_1 - U^\top U & -V_1^\top V_1 + U^\top U & -\sqrt{2}V_1^\top V_2 \\ -V_1^\top V_1 + U^\top U & 2I - V_1^\top V_1 - U^\top U & -\sqrt{2}V_1^\top V_2 \\ -\sqrt{2}V_2^\top V_1 & -\sqrt{2}V_2^\top V_1 & 2I - 2V_2^\top V_2 \end{bmatrix}.$$

Thus, $Y$ has the following structure:

$$Y = \begin{bmatrix} Y_1 & Y_2 & Y_3 \\ Y_2^\top & Y_1 & Y_3 \\ Y_3^\top & Y_3^\top & Y_4 \end{bmatrix},$$

where $Y_1^\top = Y_1, Y_2^\top = Y_2, Y_4^\top = Y_4$.
Moreover,

$$\hat{D}Y + Y\hat{D} = \begin{bmatrix} DY_1 + Y_1 D & DY_2 - Y_2 D & DY_3 \\ -DY_2^\top + Y_2^\top D & -DY_1 - Y_1 D & -DY_3 \\ Y_3^\top D & -Y_3^\top D & 0 \end{bmatrix}.$$

Hence, the structure of

$$B := \hat{V}^\top \left( \hat{M}_\tau^h(t) + \frac{1}{h}\hat{M}(t) \right) \hat{V} + \frac{1}{2}(Y\hat{D} + \hat{D}Y)$$

is as follows

$$B = \begin{bmatrix} B_1 & B_2 & B_3 \\ B_2^\top & -B_1 & -B_3 \\ B_3^\top & -B_3^\top & 0 \end{bmatrix},$$

where $B_1, B_2 \in \mathbb{R}^{n\times n}$, $B_3 \in \mathbb{R}^{n\times(m-n)}$ and $B_2^\top = -B_2$.
The non-zero entries of $Z = -Z^\top =: \begin{bmatrix} Z_1 & Z_2 & Z_3 \\ -Z_2^\top & Z_4 & Z_5 \\ -Z_3^\top & -Z_5^\top & 0 \end{bmatrix}$ are determined by

$$\hat{D}Z - Z\hat{D} = -B \Leftrightarrow$$

$$\begin{bmatrix} DZ_1 - Z_1 D & DZ_2 + Z_2 D & DZ_3 \\ DZ_2^\top + Z_2^\top D & -DZ_4 + Z_4 D & -DZ_5 \\ Z_3^\top D & -Z_5^\top D & 0 \end{bmatrix} = - \begin{bmatrix} B_1 & B_2 & B_3 \\ B_2^\top & -B_1 & -B_3 \\ B_3^\top & -B_3^\top & 0 \end{bmatrix},$$

implying that $Z_4 = Z_1$, $Z_5 = Z_3$ and $Z_2^\top = -Z_2$, since $B_2^\top = -B_2$.
Now consider

$$\frac{\sqrt{2}}{h}(\hat{V}_{k+1} - \hat{V}_k) = \frac{1}{h}\left[\begin{array}{ccc} V_1^{k+1} - V_1^k & V_1^{k+1} - V_1^k & \sqrt{2}V_2^{k+1} - \sqrt{2}V_2^k \\ U_{k+1} - U_k & -U_{k+1} + U_k & 0 \end{array}\right],$$

and the update scheme (4.52) implies that

$$\frac{1}{h}\left[\begin{array}{ccc} V_1^{k+1} - V_1^k & V_1^{k+1} - V_1^k & \sqrt{2}V_2^{k+1} - \sqrt{2}V_2^k \\ U_{k+1} - U_k & -U_{k+1} + U_k & 0 \end{array}\right] = \sqrt{2}\hat{V}_k\left(Z + \frac{1}{2}Y\right).$$

This is equivalent to

$$\frac{1}{h}\left[\begin{array}{ccc} V_1^{k+1} - V_1^k & V_1^{k+1} - V_1^k & \sqrt{2}V_2^{k+1} - \sqrt{2}V_2^k \\ U_{k+1} - U_k & -U_{k+1} + U_k & 0 \end{array}\right] =$$

$$\left[\begin{array}{c} V_1^k(Z_1 + \frac{1}{2}Y_1 - Z_2^\top + \frac{1}{2}Y_2^\top) + \sqrt{2}V_2^k(-Z_3^\top + \frac{1}{2}Y_3^\top) \\ U_k(Z_1 + \frac{1}{2}Y_1) - U_k(-Z_2^\top + \frac{1}{2}Y_2^\top) \end{array}\right.$$

$$\begin{array}{c} V_1^k(Z_2 + \frac{1}{2}Y_2 + Z_1 + \frac{1}{2}Y_1) + \sqrt{2}V_2^k(-Z_3^\top + \frac{1}{2}Y_3^\top) \\ U_k(Z_2 + \frac{1}{2}Y_2) - U_k(Z_1 + \frac{1}{2}Y_1) \end{array}$$

$$\left.\begin{array}{c} V_1^k(Z_3 + \frac{1}{2}Y_3) + V_1^k(Z_3 + \frac{1}{2}Y_3) + \sqrt{2}V_2^k(\frac{1}{2}Y_4) \\ 0 \end{array}\right],$$

from where the claimed update schemes for $V_k = [V_1^k\ V_2^k]$ and $U_k$ can be immediately seen. □

### 4.1.4 Numerical results

All simulations were performed in Matlab. We used $N = 80$ steps, step size $h = 0.05$, $A(t) = X_*(t)K(t)X_*(t)^\top$ for $K(t) = \operatorname{diag}(a_1 + \sin(t), ..., a_{10} + \sin(10t))$ and

$$X_*(t) = R^\top \left[\begin{array}{ccc} \cos(t) & \sin(t) & 0 \\ -\sin(t) & \cos(t) & 0 \\ 0 & 0 & I_8 \end{array}\right] R,$$

where $R \in O(10)$ is a fixed random orthogonal matrix and $a_i := 2.5i$ for $i = 1, ..., 10$.

**EVD with simple eigenvalues**

The definition of $A(t)$ implies the existence of a smooth curve $t \mapsto X_*(t) \in O(10)$ such that $X_*(t)A(t)X_*(t)$ is diagonal for all $t$. We wanted to determine a sequence $X_k$, which approximates $X_*(t)$ at a reasonable accuracy for $t = kh$ and $k = 1, ..., N$.

In the first simulation, we check the tracking ability of the algorithm as defined in Theorem 4.1, and we used approximations for $\dot{A}(t)$ of order 2. Figure 4.1 shows the computed (dashed) and exact (solid) time-varying eigenvalues of $A(t)$. As it can be seen in the corresponding error plot (Fig 4.2.), where $\|X_k - X_*(t_k)\|$ is depicted, we did not use perfect initial conditions. The computed values converged fast towards

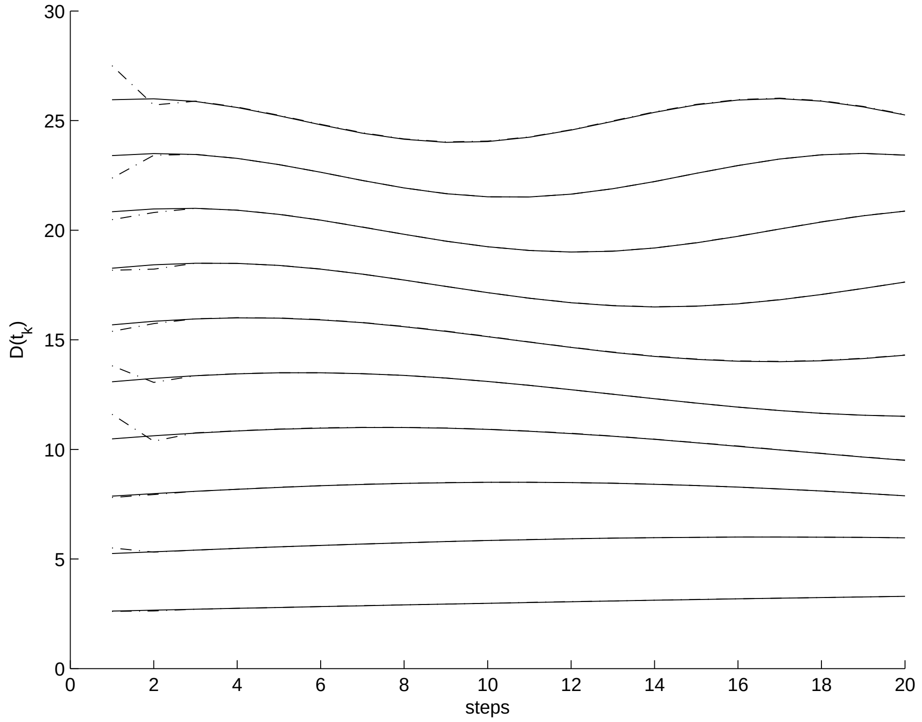


Figure 4.1: EVD with simple eigenvalues: The diagonal entries of $X_k^\top A(t_k)X_k$ (dotted) and of $X_*(t_k)^\top A(t_k)X_*(t_k)$ (solid).

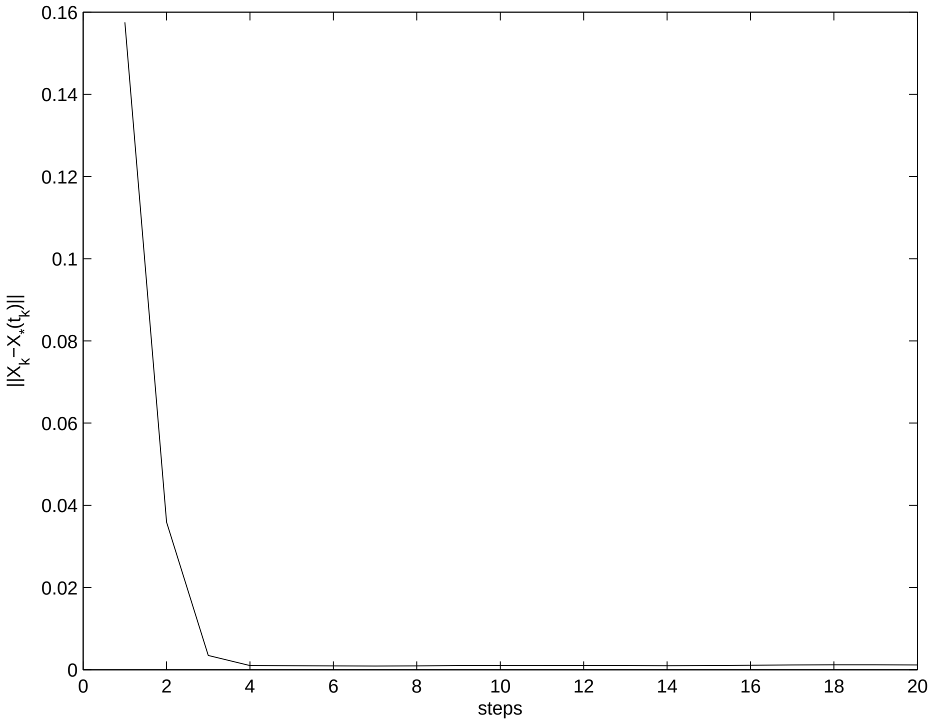


Figure 4.2: EVD with simple eigenvalues: The error plot, corresponding to Figure 4.1.

| Order $p$ | Mean error Alg. 1 (Thm 4.1) | Mean error Alg. 2 (Thm 4.2) |
|---|---|---|
| 0 | $4.5 \cdot 10^{-2}$ | $4.6 \cdot 10^{-2}$ |
| 1 | $4.1 \cdot 10^{-3}$ | $4.6 \cdot 10^{-3}$ |
| 2 | $1.4 \cdot 10^{-3}$ | $1.4 \cdot 10^{-3}$ |
| 3 | $1.2 \cdot 10^{-3}$ | $1.2 \cdot 10^{-3}$ |

Table 4.1: The mean error of the two algorithms, computed for different order approximations of $\dot{A}$

| Dimension $n \times n$ | Comp. time Alg. 1 | Comp. time Alg. 2 |
|---|---|---|
| $10 \times 10$ | $4.2 \cdot 10^{-1}$ | $1.1 \cdot 10^{-2}$ |
| $20 \times 20$ | $2.3 \cdot 10^{1}$ | $2.5 \cdot 10^{-2}$ |
| $40 \times 40$ | $9.6 \cdot 10^{2}$ | $1.4 \cdot 10^{-1}$ |

Table 4.2: The computing time of Algorithm 1 (Theorem 4.1) and Algorithm 2 (Theorem 4.2) for different sizes of $A(t) \in \mathbb{R}^{n \times n}$

the exact solution and remained close to it. Note that not all 80 computed points are shown in order to have a better resolution of the interesting initial behavior.
It is computationally much cheaper to implement the algorithm, which computes an approximation instead of the exact inverse of the Hessian $H_f$, as defined in Theorem 4.2. We therefore used this modified algorithm to do another simulation with the same initial conditions. Qualitatively, we observed the same behavior.
In order to compare this approach quantitatively with the original one from Theorem 4.1, we define the mean error by

$$\frac{1}{N} \sum_{k=1}^{N} \|X_k - X_*(t_k)\|,$$

where $N$ denotes the number of steps. Table 4.1 shows the mean accuracy of both algorithms for different choices of the order of the approximation for $\dot{A}$.
Hence, using approximations of order $p > 1$ significantly improves the quality of the results in both algorithms. Moreover, both algorithms have a comparable accuracy, which implies, that the second algorithm is preferable, since it is much cheaper to compute, cf. the computing times in Table 4.2.

**EVD with multiple eigenvalues**

In order to check the tracking property of the algorithm for symmetric matrices with multiple eigenvalues, as defined in Theorem 4.3, we slightly modified the setting from above. Instead of the above definition, we used $K(t) = \mathrm{diag}\,(a_1 + \sin(t), ..., a_8 + \sin(8t), a_8 + \sin(8t), a_8 + \sin(8t)) \in \mathbb{R}^{10 \times 10}$, $a_i := 2.5i$ for $i = 1, ..., 8$.
It turned out, that the used update scheme from Theorem 4.3 produces a sequence of matrices $(X_k)$, which approximatively diagonalizes $A(t_k)$ for $k = 1, ..., 40$. Figure 4.3 shows the diagonal entries of $X_k^\top A(t_k) X_k$, where perfect initial conditions were used.

Note that we can not visually distinguish between these computed eigenvalues and the exact ones. Only the line at the top of the panel seems to be thicker than the others. This is caused by the fact, that there are 3 almost identical lines, which correspond to the multiple eigenvalue $\lambda_8(t_k)$.

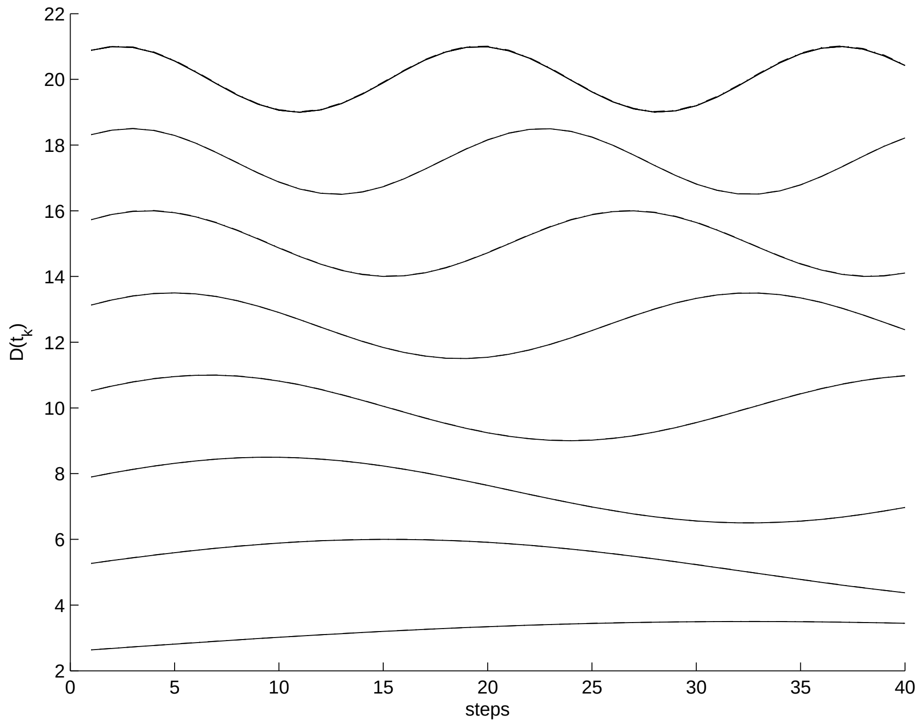


Figure 4.3: EVD with multiple eigenvalues: The diagonal entries of $X_k^\top A(t_k)X_k$ (dotted) and of $X_*(t_k)^\top A(t_k)X_*(t_k)$ (solid).

Note further that we can not just compare the output $X_k$ with $X_*(t_k)$, in order to determine the accuracy of this algorithm, as $X_*(t_k)$ is not an isolated diagonalizing transformation for $A(t_k)$. However, the first 7 columns of $X_*(t_k)$ are locally unique, and therefore, we define the mean error by

$$\frac{1}{N}\sum_{k=1}^{N}\sqrt{\|X_k^1 - X_*^1(t_k)\|^2 + \|(X_k^2)^\top X_k^2 - I_3\|^2},$$

where we used the partitions $X_k = (X_k^1\ X_k^2)$, $X_*(t_k) = (X_*^1(t_k)\ X_*^2(t_k))$ and $N$ denotes the number of steps. Table 4.3 shows the computed mean error of the algorithm from Theorem 4.3 for different order approximations of $\dot{A}$. Thus the method is able to track the desired transformation and, as expected, the error decreases for higher order approximations.

| Order $p$ | Mean error |
|---|---|
| 0 | $4.2 \cdot 10^{-2}$ |
| 1 | $3.5 \cdot 10^{-3}$ |
| 2 | $1.4 \cdot 10^{-3}$ |
| 3 | $1.2 \cdot 10^{-3}$ |

Table 4.3: The mean error of the algorithm form Theorem 4.3, computed for different order approximations of $\dot{A}$.

**SVD of a non-square matrix**

We now want to use the tracking algorithm of Theorem 4.4 to compute for a given sequence of matrices $M(t_k) \in \mathbb{R}^{m\times n}$, orthogonal matrices $U_k, V_k$, such that approximatively holds

$$V_k^\top M(t_k) U_k \approx \begin{bmatrix} \operatorname{diag}(\sigma_1(t_k), ..., \sigma_n(t_k)) \\ 0_{(m-n)\times n} \end{bmatrix}.$$

We used $m = 10$, $n = 7$ and $M(t) = V_*(t)\hat{\Sigma}U_*^\top$, where $\hat{\Sigma}(t) = \begin{bmatrix} \operatorname{diag}(\sigma_1(t), ..., \sigma_7(t)) \\ 0_{(m-n)\times n} \end{bmatrix}$, $V_*(t) = R_1^\top \begin{bmatrix} \cos(t) & \sin(t) & 0 \\ -\sin(t) & \cos(t) & 0 \\ 0 & 0 & I_8 \end{bmatrix} R_1$, $U_*(t) = R_2^\top \begin{bmatrix} \cos(t) & \sin(t) & 0 \\ -\sin(t) & \cos(t) & 0 \\ 0 & 0 & I_5 \end{bmatrix} R_2$. Here, $R_1 \in O(10)$ and $R_2 \in O(7)$ are fixed random orthogonal matrices and $\sigma_i(t) := 2.5i + \sin(it)$ for $i = 1, ..., 7$. We evaluated the algorithm at times $t_k = kh$, $k = 1, ..., N$ with $N = 20$ and $h = 0.05$.

In a first simulation, we wanted to check the tracking properties of the derived algorithm. We therefore used about 10% perturbed initial values for $V_0 \in \mathbb{R}^{10\times 10}$ and $U_0 \in \mathbb{R}^{7\times 7}$ and an approximation for $\dot{M}(t)$ of order 2. As it can be seen in Figure 4.4, the diagonal entries of $V_k^\top M(t_k)U_k$ converged fast to the exact singular values $\sigma_1(t_k), ..., \sigma_7(t_k)$, which illustrates the robustness of the used method. The corresponding error plot shows that the error initially decreased to a certain level ($\approx 2 \cdot 10^{-3}$), where it remained for the rest of the simulation, cf. Figure 4.5.

Then we wanted to examine the accuracy of the tracking algorithm for different order approximations of $\dot{M}$ and perfect initial conditions. Note that, analogously to the EVD in the case of multiple eigenvalues, the component $V_2$ of the orthogonal factor $V = [V_1\ V_2]$ is not unique, cf. Chapter 4.1.3. Therefore, we used the following formula to estimate the mean error of the algorithm

$$\frac{1}{N}\sum_{k=1}^{N} \sqrt{\|V_k^1 - V_*^1(t_k)\|^2 + \|U_k - U_*(t_k)\|^2 + \|(V_2^k)^\top V_2^k - I_{m-n}\|^2}.$$

As it can be seen in Table 4.4, this algorithm produces factors $V_k, U_k$, which perform the SVD of $M(t_k)$ at a reasonable accuracy, which increases proportionally to the order of the used approximation for $\dot{M}$.

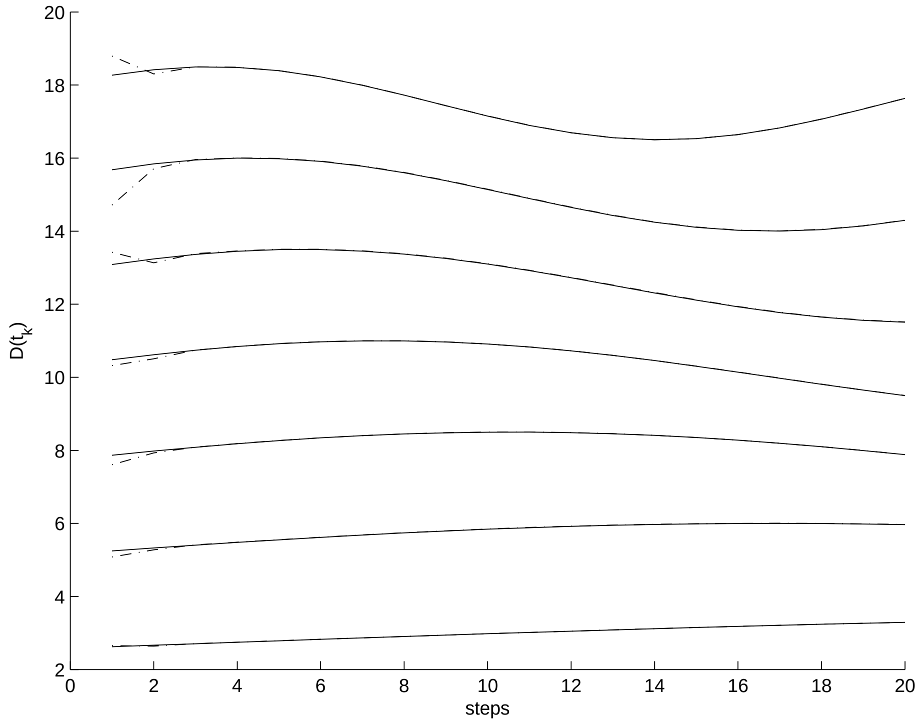


Figure 4.4: SVD: The diagonal entries of $V_k^\top M(t_k)U_k$ (dotted) and of $\hat{\Sigma}(t_k)$ (solid).

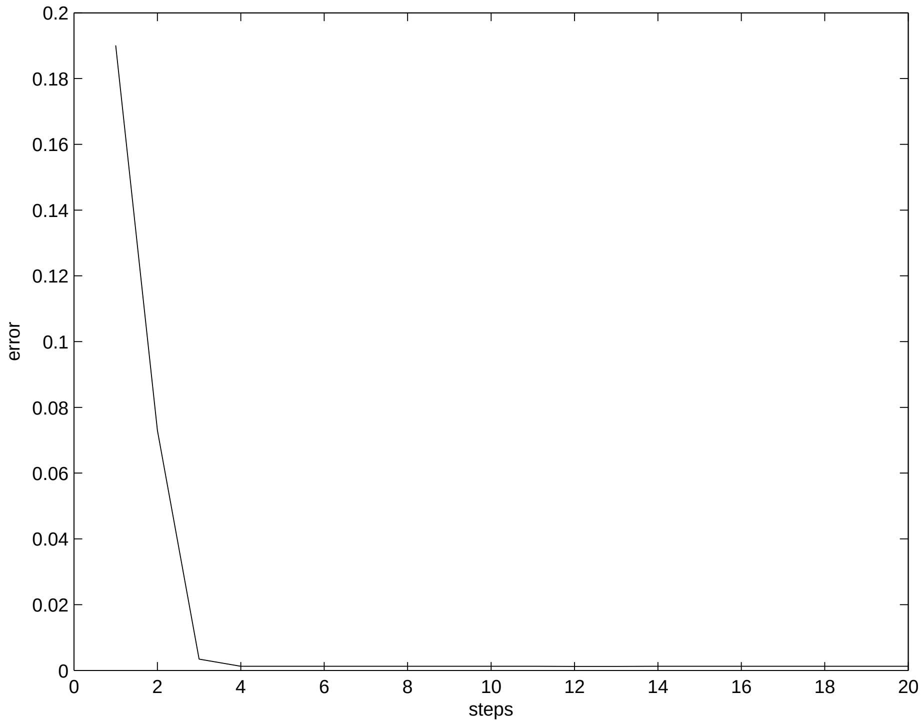


Figure 4.5: SVD: The total error of the SVD factors $U_k$, $V_k$, corresponding to Figure 4.4.

| Order $p$ | Mean error |
|---|---|
| 0 | $6.2 \cdot 10^{-2}$ |
| 1 | $4.8 \cdot 10^{-3}$ |
| 2 | $1.7 \cdot 10^{-3}$ |
| 3 | $1.5 \cdot 10^{-3}$ |

Table 4.4: The mean error of the algorithm form Theorem 4.4, computed for different order approximations of $\dot{M}$.

## 4.2 Polar decomposition of time-varying matrices

The polar decomposition of a matrix $M \in \mathbb{R}^{n\times n}$ is defined as the factorization of $M$ into a positive definite matrix $P$ and an orthogonal matrix $Z$. For full-rank matrices such factorizations always exist, which can be easily seen by considering the SVD of $M$: If $M = V\Sigma U^\top$ for some $U, V \in O(n)$ and diagonal $\Sigma \in \mathbb{R}^{n\times n}$ with positive diagonal entries, then a polar decomposition of $M$ is given by

$$M = \underbrace{V\Sigma V^\top}_{=:P}\underbrace{VU^\top}_{=:Z}. \tag{4.53}$$

Here, we study the time-varying case, i.e. we consider the polar decomposition of a family of square matrices $M(t) \in \mathbb{R}^{n\times n}$, $t \in \mathbb{R}$. At first, we introduce a SVD-based method to track the time-varying polar decomposition of $M(t)$. We then revisit an other tracking technique in order to have a benchmark for the subsequent numerical examinations.

### 4.2.1 SVD-based polar decomposition

Let $M : \mathbb{R} \to \mathbb{R}^{n\times n}$, $t \mapsto M(t)$ be a smooth curve of full rank matrices and let there exist smooth curves $t \mapsto U_*(t) \in O(n)$ and $t \mapsto V_*(t) \in O(n)$ such that

$$V_*^\top(t)M(t)U_*(t) = \Sigma(t),$$

where $\Sigma(t) = \operatorname{diag}(\sigma_1(t), ..., \sigma_n(t))$. We are able to decompose $M(t)$ into an orthogonal and a symmetric positive definite matrix as follows

$$M(t) = \underbrace{V_*(t)\Sigma(t)V_*(t)^\top}_{=:P_*(t)}\underbrace{V_*(t)U_*(t)^\top}_{=:Z_*(t)}. \tag{4.54}$$

This shows that a SVD tracking algorithm provides all necessary components in order to determine a smooth polar decomposition of $M(t)$. From Proposition 4.2, we get the following result.

**Proposition 4.3.** *Let $M(t) \in \mathbb{R}^{n\times n}$ with $\operatorname{rk}(M(t)) = n$ and singular values $\sigma_1(t), ...,$ $\sigma_n(t)$ for $t \in \mathbb{R}$. In the case, that we have groups of identical singular values, we define $\sigma_1(t) = ... = \sigma_{p_1}(t) =: \hat{\sigma}_1(t), ..., \sigma_{p_{q-1}+1}(t) = ... = \sigma_{p_q}(t) =: \hat{\sigma}_q(t)$ for $1 \leq p_1 < ... < p_q = n$ and some $q \in \mathbb{N}$.*

1. *If $t \mapsto M(t)$ is $C^r$, $r \geq 1$, and $\sigma_i(t) \neq \sigma_j(t)$ for $i \neq j$ and $t \in \mathbb{R}$, then there exists a $C^r$-polar decomposition, i.e. there exist $C^r$-families of orthogonal and positive definite matrices $Z_*(t)$, $P_*(t)$, respectively, such that $M(t) = P_*(t)Z_*(t)$.*

2. *If $t \mapsto M(t)$ is $C^r$, $r \geq 1$, and has $q$ groups of identically singular values satisfying $\hat{\sigma}_i(t) \neq \hat{\sigma}_j(t)$ for $i \neq j$ and all $t$, then there exists a $C^r$-polar decomposition of $M(t)$.*

3. *If $M(t)$ is real analytic, then there exists a real analytic polar decomposition.*

*Proof.* The claims follow directly from Proposition 4.2 by noting (4.54). □

We therefore compute the polar decomposition of $M(t)$ by applying the SVD tracking algorithm of Theorem 4.4 to the special situation here. To do this, we have to impose the following conditions on $M(t)$, cf. Section 4.1.3.

**M1** The map $t \mapsto M(t) \in \mathbb{R}^{n\times n}$, is $C^r$ with $r \geq 2$.

**M2** The singular values $\sigma_1(t), ..., \sigma_n(t)$ of $M(t)$ satisfy $\sigma_i(t) \neq \sigma_j(t)$ for $i \neq j$ and $t \in \mathbb{R}$.

**M3** $\|M(t)\|$, $\|\dot{M}(t)\|$ and $\|\ddot{M}(t)\|$ are uniformly bounded on $\mathbb{R}$.

**M4** There exists $c > 0$ such that the singular values of $M(t)$ satisfy for all $t \in \mathbb{R}$

$$|\sigma_i(t) - \sigma_j(t)| \geq c, \quad i \neq j.$$

We arrive at the following result.

**Theorem 4.5.** *Let $M(t)$ satisfy M1-M4, let $t \mapsto U_*(t) \in O(n)$ and $t \mapsto V_*(t) \in O(n)$ be $C^2$-curves satisfying $V_*^\top(t)M(t)U_*(t) = \Sigma(t)$, where $\Sigma(t) = \operatorname{diag}(\sigma_1(t), ..., \sigma_n(t))$, $t \in \mathbb{R}$. Let $M_\tau^h(t)$ denote an approximation of $\frac{\partial}{\partial t}M(t)$ of order $p \geq 1$ and let for $U, V \in \mathbb{R}^{n\times n}$, $t \in \mathbb{R}$ and $h > 0$,*

$$Y_1(V,U) = \frac{1}{2}(2I_n - V^\top V - U^\top U),$$

$$Y_2(V,U) = \frac{1}{2}(U^\top U - V^\top V),$$

$$B_1(V,U,t) =$$

$$\frac{1}{2}\left(U^\top(hM_\tau^h(t) + M(t))^\top V + V^\top(hM_\tau^h(t) + M(t))U + D(t)Y_1(V,U) + Y_1(V,U)D(t)\right),$$

$$B_2(V,U,t) =$$

$$\frac{1}{2}\left(U^\top(hM_\tau^h(t) + M(t))^\top V - V^\top(hM_\tau^h(t) + M(t))U + D(t)Y_2(V,U) - Y_2(V,U)D(t)\right),$$

*where*

$$D(V,U,t) := \operatorname{diag}\left((U^\top M(t)^\top V)_{11}, ..., (U^\top M(t)^\top V)_{nn}\right),$$

*and* $d_i(V,U,t) = (D(V,U,t))_{ii}$, $i = 1,...,n$. *Let further* $Z_1(V,U,t) = -Z_1(V,U,t)^\top \in \mathbb{R}^{n\times n}$ *and* $Z_2(V,U,t) = -Z_2(V,U,t)^\top \in \mathbb{R}^{n\times n}$ *be defined as*

$$(Z_1(V,U,t))_{ij} = \begin{cases} (B_1(V,U,t))_{ij}(d_j(V,U,t) - d_i(V,U,t))^{-1}, & i \neq j \\ 0, & \text{else} \end{cases}$$

$$(Z_2(V,U,t))_{ij} = \ -(B_2(V,U,t))_{ij}(d_j(V,U,t) + d_i(V,U,t))^{-1}, \quad i,j = 1,...,n$$

*Let* $(U_k), (V_k)$ *and* $(\Sigma)_k$ *be defined for* $k \in \mathbb{N}$ *by*

$$V_{k+1} = V_k + V_k\left(Z_1(V_k,U_k,t_k) + Z_2(V_k,U_k,t_k) + \frac{1}{2}(Y_1(V_k,U_k) + Y_2(V_k,U_k))\right),$$

$$U_{k+1} = U_k + U_k\left(Z_1(V_k,U_k,t_k) + \frac{1}{2}Y_1(V_k,U_k) - Z_2(V_k,U_k,t_k) - \frac{1}{2}Y_2(V_k,U_k)\right),$$

$$\Sigma_{k+1} := \operatorname{diag}\left((U_{k+1}^\top M(t_{k+1})^\top V_{k+1})_{11}, ..., (U_{k+1}^\top M(t_{k+1})^\top V_{k+1})_{nn}\right).$$

*Then for* $c > 0$ *and sufficiently small* $h > 0$, *the sequences* $(P_k), (Z_k)$ *defined by*

$$\boxed{\begin{aligned} P_{k+1} &= V_{k+1}\Sigma_{k+1}V_{k+1}^\top, \\ Z_{k+1} &= V_{k+1}U_{k+1}^\top, \end{aligned}} \tag{4.55}$$

*satisfy for* $k \in \mathbb{N}$ *and* $t_k = kh$

$$\|Z_k - Z_*(t_k)\| \leq ch(2 + ch)$$

*and*

$$\|P_k - P_*(t_k)\| \leq ch(2 + ch)(1 + \|\Sigma_*(t_k)\|(1 + 2ch + c^2h^2)),$$

*provided* $(V_0, U_0)$ *is sufficiently close to* $(V_*(0), U_*(0))$.

*Proof.* Under the conditions of this theorem, we have for $k \in \mathbb{N}$

$$\|V_k - V_*(t_k)\|^2 + \|U_k - U_*(t_k)\|^2 \leq c^2h^2,$$

for $(V_0, U_0)$ sufficiently close to $(V_*(0), U_*(0))$, cf. Theorem 4.4.
Thus,

$$V_kU_k^\top = V_*(t_k)U_*(t_k)^\top + \epsilon_1,$$

where $\|\epsilon_1\| \leq 2ch + c^2h^2$ and

$$\Sigma_k = \Sigma_*(t_k) + \epsilon_2',$$

where $\|\epsilon_2'\| \leq (2ch + c^2h^2)\|\Sigma_*(t_k)\|$, which shows that

$$V_k\Sigma_kV_k^\top = V_*(t_k)\Sigma_*(t_k)V_*(t_k)^\top + \epsilon_2,$$

where $\|\epsilon_2\| \leq \|\Sigma_*(t_k)\|(1 + 2ch + c^2h^2)(2ch + c^2h^2) + (2ch + c^2h^2)$. □

### 4.2.2 Polar decomposition by square root tracking

We now consider an other method to determine the time-varying polar decomposition of $M(t) \in Gl(n)$.
In the work of Getz [28], the polar decomposition of $M(t)$ is obtained by computing the uniquely determined positive square root $X_*(t)$ of $\Lambda(t) := M(t)M(t)^\top$. Then the positive definite factor $P(t)$ and the orthogonal factor $Z(t)$ such that $M(t) = P(t)Z(t)$ are given by

$$P(t) = \Lambda(t)X_*(t), \quad Z(t) = X_*(t)M(t).$$

Thus, the major problem is now the determination of the square root $X_*(t)$ of $\Lambda(t)$. This can be solved by finding the zero of the function $F : S(n) \times \mathbb{R} \to S(n)$, defined by

$$F(X,t) := X\Lambda(t)X^\top - I. \tag{4.56}$$

Here $S(n) \subset \mathbb{R}^{n\times n}$ denotes the set of all symmetric matrices. Obviously, $\dim S(n) = \frac{n(n+1)}{2} =: s(n)$.
The zero of $F$ can be determined by using the time-varying Newton flow, i.e. by finding a solution $X(t)$ of

$$DF(X,t)\dot{X} = -\mu F(X,t) - \frac{\partial}{\partial t}F(X,t), \quad \mu > 0.$$

One therefore needs to invert $DF(X(t),t) \in \mathbb{R}^{s(n)\times s(n)}$ to get an explicit version of the above differential equation. According to Getz [28], the solution $\Gamma(t)$ of the differential equation

$$\dot{\Gamma} = -\mu DF(X_*(t),t)^\top(DF(X_*(t),t)\Gamma - I) - \Gamma\frac{\partial}{\partial t}DF(X_*(t),t)\Gamma$$

converges exponentially to $DF(X_*(t),t)^{-1}$, provided that $\mu > 0$ is sufficiently large and $\Gamma(0)$ is sufficiently close to $DF(X_*(0),0)^{-1}$
Thus, Getz proposes to solve the following coupled system to compute $X(t) \approx X_*(t)$

$$\left\{\begin{array}{lcl} \operatorname{vec}(\dot{X}) & = & -\mu\Gamma\operatorname{vec}(X\Lambda(t)X - I) - \Gamma\operatorname{vec}(X\dot{\Lambda}X) \\ \dot{\Gamma} & = & -\mu\Gamma(DF(X,t)\Gamma - I) - \Gamma(\frac{d}{dt}DF(X,t))\big|_{\dot{X}=-\Gamma\operatorname{vec}(X\dot{\Lambda}(t)X)}\Gamma \end{array}\right. \tag{4.57}$$

Here $\operatorname{vec} : S(n) \to \mathbb{R}^{s(n)}$ denotes a vectorizing operation to transform symmetric matrices into equivalent vectors in $\mathbb{R}^{s(n)}$.
In particular, one has the following theorem ([28]).

**Theorem 4.6.** *Let $M(t)$ be in $Gl(n)$ for all $t \in \mathbb{R}$. Let the polar decomposition of $M(t)$ be $M(t) = P(t)Z(t)$ with $P(t) \in S(n)$ the positive definite symmetric square root of $\Lambda(t) := M(t)M(t)^\top$ and $Z(t) \in O(n)$ for all $t \in \mathbb{R}$. Let $\operatorname{vec}(X) \in \mathbb{R}^{s(n)}$, $\Gamma \in \mathbb{R}^{s(n)\times s(n)}$ and let $(\operatorname{vec}(X(t)), \Gamma(t))$ denote the solution of (4.57), where $F$ is defined by (4.56). Then for $\mu > 0$ sufficiently large and $(\operatorname{vec}(X(0)), \Gamma(0))$ sufficiently close to $(\operatorname{vec}(X_*(0)), DF(X_*(0),0)^{-1})$ the following statements hold:*

1. *$X(t)\Lambda(t)$ exponentially converges to $P(t)$,*
2. *$X(t)M(t)$ exponentially converges to $Z(t)$,*
3. *$M(t)^\top X(t)^2$ exponentially converges to $M(t)^{-1}$.*

**Discrete tracking method**

Since Getz' approach bases on the continuous version of the Euclidean time-varying Newton flow, we use the corresponding discrete tracking algorithm (2.59) to determine $P_k$ and $Z_k$, approximating the polar decomposition of $M(t) = P_*(t)Z_*(t)$ at discrete times $t_k = kh$ for $k \in \mathbb{N}$ and step size $h > 0$. Thus for $(P_0, Z_0)$ sufficiently close to $(P_*(0), Z_*(0))$, the sequences $(P_k)$ and $(Z_k)$ are recursively given by

$$\boxed{\begin{aligned} P_{k+1} &= X_{k+1}\Lambda(t_{k+1}) \\ Z_{k+1} &= X_{k+1}M(t_{k+1}), \end{aligned}} \tag{4.58}$$

where $X_k$ is obtained by discretizing the ODE (4.57) for $\mu := \frac{1}{h}$. Hence,

$$\begin{cases} \mathrm{vec}(X_{k+1}) = \mathrm{vec}(X_k) - \Gamma_k \mathrm{vec}(X_k\Lambda(t_k)X_k - I) - h\Gamma_k \mathrm{vec}(X_k\Lambda_\tau^h(t_k)X_k) \\ \Gamma_{k+1} = \Gamma_k - \Gamma_k(DF(X_k, t_k)\Gamma_k - I) - h\Gamma_k(\frac{d}{dt}DF(X_k, t_k))\big|_{\dot{X}_k = -\Gamma_k \mathrm{vec}(X_k\Lambda_\tau^h(t_k)X_k)}\Gamma_k \end{cases}$$

Here, $\Lambda_\tau^h(t)$ denotes an approximation of $\dot{\Lambda}(t)$.
In the next section, we will compare the results of this tracking algorithm with that one defined in Theorem 4.5.

### 4.2.3 Numerical results

At first, we computed the polar decomposition of a sequence of matrices $M(t_k) \in \mathbb{R}^{n\times n}$ by using the algorithm of Theorem 4.5. To be able to compare the computed values with the exact ones, we defined $(M_k)$ in the following way.
We set $n = 10$, $\Sigma(t) = \mathrm{diag}\,(\sigma_1(t), ..., \sigma_{10}(t))$, where $\sigma_i(t) := 2.5i + \sin(it)$ for $i = 1, ..., 10$ and $t_k := kh$ for $k = 1, ..., N$. Here $h := 0.05$ is the step size and $N = 40$ is the number of steps.
Then $M(t_k)$ is given by

$$M(t_k) = V_*(t_k)\Sigma(t_k)U_*(t_k),$$

where

$$V_*(t) = R_1^\top \begin{bmatrix} \cos(t) & \sin(t) & 0 \\ -\sin(t) & \cos(t) & 0 \\ 0 & 0 & I_8 \end{bmatrix} R_1,\ U_*(t) = R_2^\top \begin{bmatrix} \cos(t) & \sin(t) & 0 \\ -\sin(t) & \cos(t) & 0 \\ 0 & 0 & I_8 \end{bmatrix} R_2$$

for fixed random orthogonal matrices $R_1, R_2 \in O(10)$. In all simulations, we used a second order approximation $M_\tau^h$ for $\dot{M}$.

Since our algorithm works by determining the SVD of $M(t)$, we used about 10% perturbed initial values $V_0, U_0$ to demonstrate the robustness of the used method. Figure 4.6 reflects the error plots of this simulation, where we observe an initial convergence of the errors to zero, where they remain at a small level ($\leq 0.08$).
Note that the error of the orthogonal factor $\|Z_k - Z_*(t_k)\| = \|V_k U_k^\top - V_*(t_k) U_*(t_k)^\top\|$ is considerably smaller than the error of the positive definite factor $\|P_k - P_*(t_k)\| = \|V_k D_k V_k^\top - V_*(t_k)\Sigma(t_k)V_*(t_k)^\top\|$. This is as expected from the error estimates in Theorem 4.5, where the error of the positive definite factor is about $\|\Sigma_*(t_k)\|$ times higher.

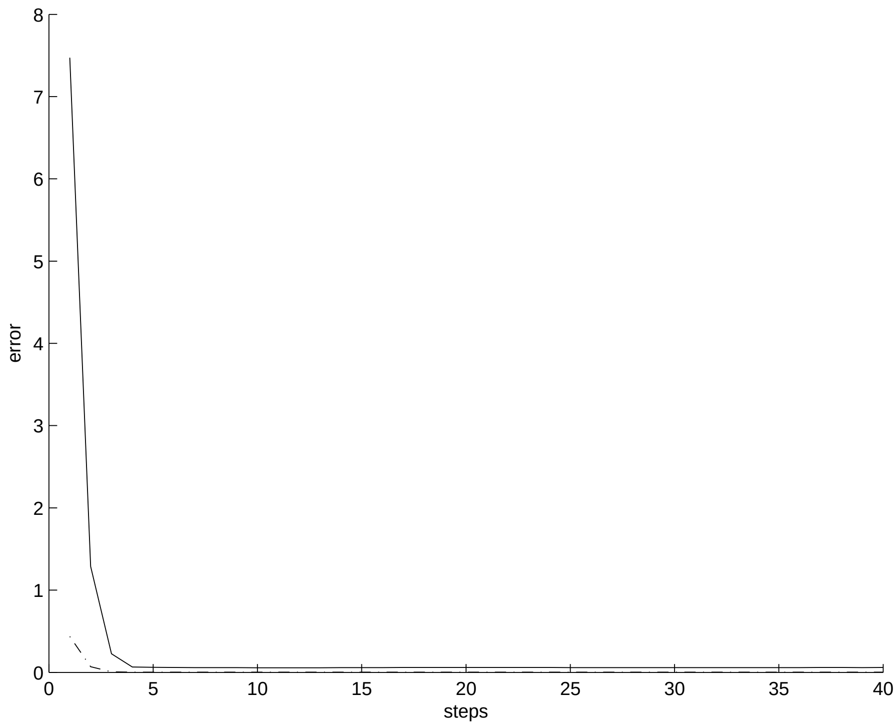


Figure 4.6: The errors of the factors of the polar decomposition of $M_k$: $\|P_k - P_*(t_k)\|$ and $\|Z_k - Z_*(t_k)\|$ (dashed). Our algorithm of Theorem 4.5 was used to compute $P_k$ and $Z_k$.

Next, we wanted to evaluate the second algorithm (Getz), as defined in (4.58). We used the same setup as above and computed the factors $P_k, Z_k$ of the polar decomposition of $M_k$. To check the robustness of this method, we used a 10% perturbed initial value for $X_0$.
Note that in order to implement this algorithm, one particularly needs to vectorize symmetric matrices and use the corresponding matrix representation of $DF$. For example, we can use the standard vectorizing operation for matrices in $\mathbb{R}^{n\times n}$. Then the matrix representation $H \in \mathbb{R}^{n^2\times n^2}$ of $DF$ can be easily determined, and is given by

$$H(X_k, t_k) = (X_k\Lambda(t_k)) \otimes I + I \otimes (X_k\Lambda(t_k)),$$

where $\otimes$ denotes the Kronecker product.

If one chooses a vectorizing operation without redundancies, one obtains a matrix representation of $DF$ of dimension $\frac{n(n+1)}{2} \times \frac{n(n+1)}{2}$, which possibly leads to a faster working algorithm. However, since the accuracy of such an algorithm would not be improved, our particular choice of the vectorizing operation is valid to check the tracking properties of this approach.
The resulting magnitudes $P_k, Z_k$ of this algorithm are compared to the exact ones $P_*(t_k)$ and $Z_*(t_k)$, which is shown by the error plot in Figure 4.7. Similarly to the first algorithm of Theorem 4.5, one also observes a fast initial convergence of the errors towards zero, where they remain for the rest of the simulation. Although we used comparable initial conditions to the first algorithm, the final error level is about twice as high ($\leq 0.17$). An other advantage of the first algorithm is, that it works with matrices of dimension $n \times n$ instead of (at least) $\frac{n(n+1)}{2} \times \frac{n(n+1)}{2}$.

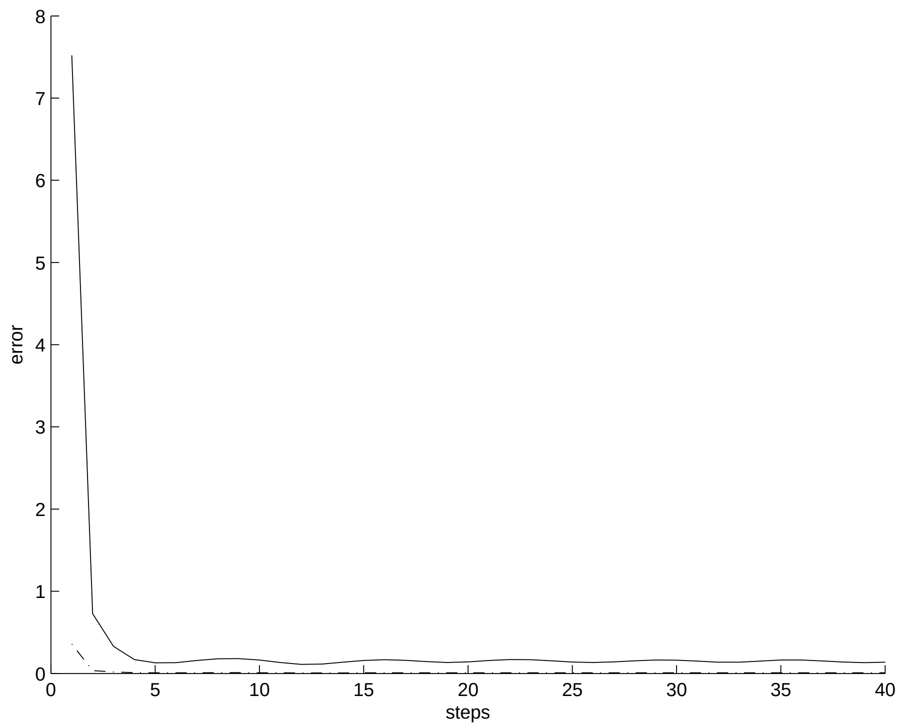


Figure 4.7: The errors of the factors of the polar decomposition of $M_k$: $\|P_k - P_*(t_k)\|$ and $\|Z_k - Z_*(t_k)\|$ (dashed). The discrete version of Getz' algorithm, defined in (4.58), was used to compute $P_k$ and $Z_k$.

## 4.3 Minor and principal eigenvector tracking

Many problems in control and signal processing require the tracking of certain eigenvectors of a time-varying matrix; the eigenvectors associated with the largest eigenvalues are called the principal eigenvectors and those with the smallest eigenvalues the minor eigenvectors.

In statistical analysis, the principal components of a covariance matrix $C$ are projections of the data vectors on the directions of the principal eigenvectors of $C$. Therefore, the major task in PCA is the determination of the principal eigenvectors of the covariance matrix.
Here, we consider the more general task of determining minor or principal eigenvectors. One interesting feature, inherited from a recently proposed minor eigenvector flow upon which part of this work is based, is that the algorithm can be used also for tracking principal eigenvectors simply by changing the sign of the matrix whose eigenvectors are being tracked. The other key feature is that the algorithm has a guaranteed accuracy, which bases on the particular choice of the tracking method.

We now consider the task of determining for $t \in \mathbb{R}$ the minor and principal eigenvectors of a time-varying symmetric matrix $A(t) \in \mathbb{R}^{n \times n}$. These are defined as the $n \times p$-matrix consisting of eigenvectors of $A(t)$ associated with the $p$ smallest or largest eigenvalues, respectively.
In the paper of Manton et al. [48] a method was introduced, which was able to extract the minor and principal eigenvectors of a constant matrix $A \in \mathbb{R}^{n \times n}$. This was achieved by finding the minimum or maximum of a suitable cost function. The minimization of this function leads to a method, which determines the minor eigenvectors, hence the eigenvectors associated to the smallest eigenvalues. We follow this approach in order to derive a minor/principal eigenvector flow for time-varying matrices $A(t)$.
In principal component analysis for time-varying data one is concerned with the associated eigenvector estimation task for a time-varying covariance matrix $A(t)$. There are at least two interpretations.

1. Let $x(\tau) \in \mathbb{R}^n$ a curve of data points for $\tau \geq 0$. Then define for $t \geq 0$

$$A(t) := \frac{1}{t} \int_0^t x(\tau) x(\tau)^\top d\tau.$$

2. Let $x_1(t), ..., x_m(t) \in \mathbb{R}^n$ curves of data points for $t \geq 0$. Then

$$A(t) := \frac{1}{m} \sum_{i=1}^{m} x_i(t) x(t)^\top.$$

The subsequent analysis will not depend on any such statistical interpretation of $A(t)$. Thus, the developed eigenvector tracking techniques are also suitable to compute the time-varying principal components of any covariance matrix $A(t)$.

Consider the smooth cost function $f : \mathbb{R}^{n \times p} \times \mathbb{R} \to \mathbb{R}$

$$f(X, t) = \frac{1}{2} \mathrm{tr}\,(A(t) X N X^\top) + \frac{\mu}{4} \|N - X^\top X\|^2, \tag{4.59}$$

where $N \in \mathbb{R}^{p \times p}$, $\mu \in \mathbb{R}$. To ensure that the extrema $X_*(t)$ of $f$ correspond to the minor/principal eigenvectors of $A(t)$, we make the following assumptions, cf. [48].

**A1** The scalar $\mu$ is strictly positive.

**A2** $A(t) \in \mathbb{R}^{n\times n}$ is symmetric for all $t \in \mathbb{R}$.

**A3** $N = \operatorname{diag}(n_1, ..., n_p) \in \mathbb{R}^{p\times p}$ with $n_1 > ... > n_p > 0$.

**A4** $A(t)$ has exactly $p$ minor (principal) eigenvectors and the corresponding eigenvalues are pairwise distinct for all $t \in \mathbb{R}$.

**A5** $\mu$ does not equal any eigenvalue of $A(t)$ for all $t \in \mathbb{R}$.

Assumption A1 implies that the cost function (4.59) has compact sublevel sets and therefore a global minimum $X_*(t)$ exists for any $t \in \mathbb{R}$. Furthermore, for fixed $t \in \mathbb{R}$, assumptions A1-A3 imply that each column of a critical point $X_* = [x_1...x_p]$ of $f$ is either the null-vector or an eigenvector of $A(t)$ with eigenvalue $\lambda_i$, $i = 1, ..., p$. The following lemma holds.

**Lemma 4.5.** *[48] Assume A1-A5 hold. For $t \in \mathbb{R}$, let $\lambda_1(t) < ... < \lambda_p(t) < \lambda_{p+1}(t) \leq \cdots \leq \lambda_n(t)$ be the eigenvalues of $A(t)$ in ascending order and let $v_1(t), ..., v_n(t)$ be the corresponding normalized eigenvectors. Then for any $t \in \mathbb{R}$, $X_*(t) = [x_1(t)\ ...\ x_p(t)]$ is a local minimum of $f(X, t)$ if and only if*

$$x_i(t) = \pm\gamma_i(t)v_i(t) \tag{4.60}$$

*where*

$$\gamma_i(t) = \begin{cases} \sqrt{N_{ii}(1 - \lambda_i(t)/\mu)} & \text{if } \lambda_i(t) < \mu \\ 0 & \text{otherwise} \end{cases}$$

*for $i = 1, ..., p$.*

Let $M$ denote an upper bound for $\lambda_i(t)$ for all $t \in \mathbb{R}$ and $i = 1, ..., p$. If $\mu > M$, then the previous lemma shows, that for any $t \in \mathbb{R}$, $X_* = [x_1\ ...\ x_p]$ is a minimum of $f(X, t)$ if and only if its columns are the non-trivial eigenvectors of $A(t)$, which correspond to the $p$ smallest eigenvalues. Thus the global minimum of $f$ gives the minor eigenvectors of $A$. In contrast, the maxima of $f$ correspond to the principal eigenvectors of $A$, i.e. the minima of the cost function $f_-$ obtained by replacing $A$ by $-A$. Thus, by replacing $A$ by $-A$ in the subsequent formulas, all results about minor eigenvectors are immediately reformulated into equivalent results about principal eigenvectors. Thus, from now on, we restrict ourselves to the minor eigenvector case. This duality between minor and principal eigenvectors does not hold for the previously proposed cost functions for principal eigenvector analysis and motivates our choice of the specific cost function (4.59).

**Proposition 4.4.** *Let $t \mapsto A(t)$ be a smooth matrix valued map satisfying A1-A5. Assume that $\|A(t)\|$, $\|\dot{A}(t)\|$ and $|\lambda_i(t) - \lambda_j(t)|^{-1}$ for $i \neq j$ are uniformly bounded on $\mathbb{R}$. Then there exists a smooth map $t \mapsto X_*(t) \in \mathbb{R}^{n\times p}$, which is a smooth isolated minimum of $f(X, t)$. Moreover, if $\mu > M$, then the rows of $X_*(t)$ are the $p$ non-trivial eigenvectors, associated to the $p$ smallest eigenvalues.*

*Proof.* Under the conditions above, there exists a smooth map $t \mapsto V(t) \in St(n,p)$ such that $V(t)^\top A(t)V(t)$ is a diagonal matrix with diagonal entries $\lambda_1(t) < ... < \lambda_p(t)$, cf. [41]. Here $St(n,p)$ denotes the Stiefel manifold of orthogonal $n \times p$ matrices $V$. Let the columns $v_1(t), ..., v_p(t)$ of $V(t)$ be ordered, such that they correspond to the eigenvalues $\lambda_1(t) < ... < \lambda_p(t)$.
Then define $x_1(t), ..., x_p(t)$ as in (4.60). Thus $X_* := [x_1(t) \; ... \; x_p(t)]$ is a critical point of $f(X,t)$. If for some $k \in \{1, ..., p\}$ and some $t \in \mathbb{R}$

$$\mu < \lambda_k(t),$$

then this relation holds for all $t \in \mathbb{R}$, due to A5. Hence, the corresponding row $x_k(t) = 0$ for all $t$. This proves the claim. □

**Remark 4.1.** The previous proof shows also, that the minimum of $f(X,t)$ is unique, except for arbitrary time-varying signs and possible permutations of its columns.

### 4.3.1 Minor eigenvector tracking

Let $A(t)$ and $f$ as above and let $X_*(t)$ denote a time-varying minimum of $f$, i.e. $f(X_*(t),t)$ is minimal for all $t$. We want to track this minimum by using the time-varying Newton flow. Hence, we need to determine the "$X$-gradient" and "$X$-Hessian" of $f$. Thus we consider for fixed $t$ the function $X \mapsto f(X,t)$ and determine its gradient, which is given as

$$\nabla f(X,t) = A(t)XN - \mu XN + \mu XX^\top X. \tag{4.61}$$

Moreover, the Hessian (with respect to $X$) reads

$$H_f(X,t) \cdot \xi = A(t)\xi N - \mu \xi N + \mu(\xi X^\top X + X\xi^\top X + XX^\top \xi). \tag{4.62}$$

In order to use the Newton flow for the tracking of the zero of the gradient, we have to compute an inverse of the Hessian operator.
In the special case $p = 1$, $X$ and $\xi$ are vectors of length $n$ and (4.62) can be rewritten as

$$H_f(X,t) \cdot \xi = A(t)\xi - \mu\xi + \mu(X^\top X + 2XX^\top)\xi, \tag{4.63}$$

where we choose $N = 1$. Hence, the matrix representation of the Hessian is just given as $A(t) - \mu I_n + \mu(X^\top X + 2XX^\top)$.

But if $p > 1$, one needs more effort to determine the Hessian operator and we show two different approaches.

#### 4.3.1.1 Vectorizing the matrix differential equation

In this paragraph, we vectorize $\xi$ and $\nabla f(X,t)$ by using the VEC operation and computing a matrix representation for $H_f$ by employing the Kronecker product, denoted by $\otimes$, cf. appendix. Thus

$$H_f(X,t) = N \otimes (A(t) - \mu I_n) + \mu(X^\top X \otimes I_n + (X^\top \otimes X)\pi_T + I_p \otimes XX^\top), \tag{4.64}$$

where $\pi_T$ is such that vectorized matrices are mapped onto vectors, which equal the vectorized transposed matrix, i.e.

$$\pi_T \cdot \mathrm{VEC}(X) = \mathrm{VEC}(X^\top), \quad X \in \mathbb{R}^{n\times p}.$$

Therefore, $\pi_T$ is a permutation matrix, which is for $p = 1$ given as $\pi_T = I_n$.

Once having determined the Hessian, we arrive at the tracking algorithm for vector-valued $X_k$, $k \in \mathbb{N}$.

**Algorithm 1:** Choose a starting point $X_0$ close to the exact minimum $X_*(0)$ of $f(X, 0)$, $t_k := kh$. For $k \in \mathbb{N}$, the new point $X_{k+1}$, which approximates the minimum of $f$ at $t_{k+1}$ is determined by

$$\boxed{X_{k+1} = X_k - H_f^{-1}\left(\nabla f(X_k, t_k) + hG_h(X_k, t_k)\right)} \tag{4.65}$$

Here $G_h$ denotes an approximation to $\frac{\partial}{\partial t}\nabla f(X_k, t_k)$ of order $m \geq 1$. Some valid choices for this are given in Section 2.1.2.2. The following theorem gives necessary conditions for the applicability of this algorithm.

**Theorem 4.7.** *Let $f(X, t)$ as defined in (4.59) and assume A1-A5. Let further $t \mapsto A(t)$ be a smooth map and let $\|A(t)\|$, $\|\dot{A}(t)\|$, $\|\ddot{A}(t)\|$, $|\lambda_i(t) - \lambda_j(t)|^{-1}$ be uniformly bounded on $\mathbb{R}$ for $i \neq j$ and $i \leq p \vee j \leq p$ and let $\mu > \sup\{\lambda_i(t) |\ 1 \leq i \leq p,\ t \in \mathbb{R}\}$.*
*Then there exists a smooth isolated zero $X_*(t)$ of $\nabla f(X, t)$, whose columns are the eigenvectors of $A(t)$, associated with the $p$ smallest values of the eigenvalues.*
*Moreover, for any sufficiently small step size $h$ and $c > 0$ with $\|X_0 - X_*(0)\| \leq ch$, the sequence $X_k$ defined by Algorithm 1 satisfies for all $k \in \mathbb{N}$*

$$\|X_k - X_*(t_k)\| \leq ch. \tag{4.66}$$

*Proof.* The existence of such a zero $X_*(t)$ is clear due to Proposition 4.4.
To show (4.66), we have to check the necessary conditions of Theorem 2.10. Therefore, we consider the derivatives of $\nabla f(X, t) = A(t)XN - \mu XN + \mu XX^\top X$:

$$\frac{\partial}{\partial t}\nabla f(X, t) = \dot{A}(t)XN$$

$$H_f(X, t) \cdot \xi = A(t)\xi N - \mu\xi N + \mu(\xi X^\top X + X\xi^\top X + XX^\top \xi)$$

$$\frac{\partial^2}{\partial t^2}\nabla f(X, t) = \ddot{A}(t)XN$$

$$D^2\nabla f(X, t) \cdot (\xi, \eta) = \mu(\xi\eta^\top X + \xi X^\top \eta + \eta\xi^\top X + X\xi^\top \eta + \eta X^\top \xi + X\eta^\top \xi)$$

$$\frac{\partial}{\partial t} H_f(X,t) \cdot \xi = \dot{A}(t)\xi N$$

Lemma 4.5 shows that $\sup_{t\in\mathbb{R}} \|X_*(t)\|$ is finite. This shows, that for some $r > 0$ these derivatives are bounded for all $t \in \mathbb{R}$ and $X \in B_r(X_*(t))$.
The only estimate which remains to show is, that there exists a $M > 0$ such that for all $t \in \mathbb{R}$

$$\|H_f(X_*(t),t)^{-1}\| \leq M \tag{4.67}$$

According to [48], the eigenvalues of the Hessian at a critical point $X_*(t)$ are

$$\{2N_{ii}(\mu - \lambda_i(t)),\ i = 1,...,p\} \cup \{N_{ii}(\lambda_j(t) - \lambda_i(t)),\ i = 1,...,p,\ j = p+1,...,n\} \cup$$
$$\{(N_{ii} - N_{jj})(\lambda_j(t) - \lambda_i(t)),\ 1 \leq i < j \leq p\} \cup$$
$$\{N_{ii}(\mu - \lambda_i(t)) + N_{jj}(\mu - \lambda_j(t)),\ 1 \leq i < j \leq p\}.$$

As we assumed the distances $|\lambda_j(t) - \lambda_i(t)|$, $1 \leq i \neq j \leq n$ to be uniformly bounded in $t$, the smallest absolute value of the eigenvalues of $H_f(X_*(t),t)$ is lower bounded for all $t$, which shows (4.67). □

Note that the matrix representation of $H_f(X,t)$ is of dimension $np \times np$, which shows, that standard way to employ the Newton flow by solving a linear system

$$H_f(X,t)\dot{X} = -\nabla f(X,t) + hG_h(X,t)$$

for $\dot{X}$ might, however, not be practical for large $n$ and $p$. This motivates to look for an other way to invert $H_f$.

#### 4.3.1.2 Approximately solving the implicit differential equation

In order to derive a practical inversion formula for $H_f$,we consider the following equation for $\xi$

$$H_f(X,t)\xi := A(t)\xi N - \mu\xi N + \mu(\xi X^\top X + X\xi^\top X + XX^\top \xi) = R, \tag{4.68}$$

where $R \in \mathbb{R}^{n\times p}$. If $X_*(t)$ denotes a minimum of $f$, we determine $Q \in O(n)$ such that $QD = X_*$, where $D \in \mathbb{R}^{n\times p}$ satisfying $D_{ij} = 0$ for $i \neq j$. Such matrices $Q, D$ exist, as the columns of $X_*$ are pairwise orthogonal. Hence, $D_{ii} = \pm\|x_i\|$, and $q_i = x_i/\|x_i\|$ where $x_i, q_i$ denotes the $i$th column of $X_*, Q$, respectively for $i = 1,...,p$. The remaining $n-p$ columns $Y$ of $Q$ have to be chosen such that $Q$ is an orthogonal matrix, and the columns of $Y$ span the eigenspace of the $n-p$ principal eigenvectors.
This is a quite restrictive assumption, however, if the $n-p$ principal eigenvectors are all equal, then $Y \in \mathbb{R}^{n\times(n-p)}$ can be any matrix with orthonormal columns such that $X^\top Y = 0$. Note further, that in the case of $p = n$, this assumption is not a restriction, either.

By defining $\eta := Q^\top \xi$, we can rewrite the above equation in $X = X_*(t) = QD$ as

$$(A(t) - \mu I_n)Q\eta N + \mu(Q\eta D^\top Q^\top QD + QD\eta^\top Q^\top QD + QDD^\top Q^\top Q\eta) = R \quad (4.69)$$

Multiply both sides with $Q^\top$. Then

$$Q^\top (A(t) - \mu I_n)Q\eta N + \mu Q^\top (Q\eta D^\top D + QD\eta^\top D + QDD^\top \eta) = Q^\top R, \quad (4.70)$$

which is for orthogonal $Q$ and diagonal $K = Q^\top (A(t) - \mu I_n)Q$ equivalent to

$$K\eta N + \mu(\eta D^\top D + D\eta^\top D + DD^\top \eta) = Q^\top R \quad (4.71)$$

This is a linear equation which we want to solve for $\eta$. As we have assumed that $X_*$ is a minimum of $f(X,t)$, the matrix $K = Q^\top (A(t) - \mu I_n)Q = Q^\top A(t)Q - \mu I_n$ is diagonal with distinct eigenvalues. Thus, we can solve equation (4.71) by considering the entries on position $(i,j)$ and $(j,i)$ for $1 \leq i,j \leq p$. We have

$$c_1\eta_{ij} + c_2\eta_{ji} = (Q^\top R)_{ij}$$

and

$$c_3\eta_{ij} + c_4\eta_{ji} = (Q^\top R)_{ji}$$

Here $c_1 = K_{ii}N_{jj} + \mu(D_{jj}^2 + D_{ii}^2)$, $c_2 = D_{ii}D_{jj}$, $c_3 = K_{jj}N_{ii} + \mu(D_{ii}^2 + D_{jj}^2)$, $c_4 = D_{jj}D_{ii}$. If $i > p$ then $D_{ij} = 0$ for all $j = 1, ..., p$. Hence, $(D\eta^\top D)_{ij} = 0$ and we have only to consider one equation to determine $\eta_{ij}$:

$$(c_1 + c_5)\eta_{ij} = (Q^\top R)_{ij},$$

where $c_5 = \mu D_{jj}^2$. Therefore, we get the following formula for $\eta$:

$$\eta_{ij} = \begin{cases} \left((Q^\top R)_{ij} - \frac{c_2}{c_3}(Q^\top R)_{ji}\right)\left(c_1 - \frac{c_2c_4}{c_3}\right)^{-1}, & \text{for } 1 \leq i,j \leq p \\ \frac{(Q^\top R)_{ij}}{c_1 + c_5}, & \text{for } 1 \leq j \leq p,\ p+1 \leq i \leq n \end{cases} \quad (4.72)$$

Noting that $\xi = Q\eta$ shows that we have now found an explicit form of (4.68) in $X = X_*(t)$.

If $X$ is not the exact minimum of $f$, then $Q = [XD \quad Y]$ is not orthogonal, where $D, Y$ are chosen as described above. Moreover, $Q^\top (A(t) - \mu I_n)Q$ is not diagonal and hence, $K := \operatorname{diag}(Q^\top (A(t) - \mu I_n)Q)$ is only approximation for that matrix. However, by using this simplification, one can approximatively solve (4.70) by solving equation (4.71), whose solutions in $\eta$ are given by (4.72). Since the original system (4.70) is well-conditioned in $X = X_*(t)$, these approximative solutions are expedient.
We arrive at the following tracking algorithm. Note that the implementation of this algorithm is considerably cheaper than the previous one, as there is no need to compute the exact inverse of the Hessian.

**Algorithm 2:**
1. Choose a starting point $X_0$ close to the exact minimum $X_*(0)$ of $f(X,0)$ and use a sufficiently small step size $h$ and $t_k := kh$.
2. For $k \in \mathbb{N}$, suppose that $X_k$ is given. Choose a matrix $Q_k \in \mathbb{R}^{n\times n}$, whose first $p$ columns result from the normalization of the columns of $X_k$. The remaining $n-p$ columns $Y$ of $Q$ have to be chosen such that they are normalized and approximatively span the eigenspace $Y_* \in \mathbb{R}^{n\times(n-p)}$ of the $n-p$ principal eigenvectors, i.e. there exists a $k>0$ such that if $X = X_* + \epsilon$, then $\|Y - Y_*\| \leq k\|\epsilon\|$.
3. The new point $X_{k+1}$, which approximates the minimum of $f$ at $t_{k+1}$ is defined by

$$\boxed{X_{k+1} = X_k - Q_k\eta,} \tag{4.73}$$

where $\eta \in \mathbb{R}^{n\times p}$ is as defined in (4.72) for $R = \nabla f(X_k, t_k) + hG_h(X_k, t_k)$ and $G_h$ denotes the approximation to $\frac{\partial}{\partial t}\nabla f(X_k, t_k)$ of order $m \geq 1$.

**Theorem 4.8.** *Let $f(X,t)$ as defined in (4.59) and assume A1-A5. Let further $t \mapsto A(t)$ be a smooth map and let $\|A(t)\|$, $\|\dot{A}(t)\|$, $\|\ddot{A}(t)\|$, $|\lambda_i(t) - \lambda_j(t)|^{-1}$ be uniformly bounded on $\mathbb{R}$ for $i \neq j$ and $i \leq p \vee j \leq p$ and let $\mu > \sup\{\lambda_i(t)|\ 1 \leq i \leq p,\ t \in \mathbb{R}\}$.*
*Then there exists a smooth isolated zero $X_*(t)$ of $\nabla f(X,t)$, whose columns are the eigenvectors of $A(t)$, associated with the $p$ smallest values of the eigenvalues.*
*Moreover, for any sufficiently small step size $h$ and $c > 0$ with $\|X_0 - X_*(0)\| \leq ch$, the sequence $X_k$ defined by Algorithm 2 satisfies for all $k \in \mathbb{N}$*

$$\|X_k - X_*(t_k)\| \leq ch. \tag{4.74}$$

*Proof.* The existence of such a zero $X_*(t)$ is clear due to Proposition 4.4. To show (4.74), we have to check the conditions of Theorem 2.12.
As the boundedness properties of $\|H_f(x_*(t),t)\|$, $\|\frac{\partial}{\partial t}\nabla f(x_*(t),t)\|$, $\|H_f(x_*(t),t)^{-1}\|$, $\|DH_f(x,t)\|$ and $\|\frac{\partial^2}{\partial t^2}\nabla f(x,t)\|$ already have been shown in the proof of Theorem 4.7, we only show the additional assumptions.
The fact, that $\|\frac{\partial}{\partial t}H_f(x,t)\|$ is uniformly bounded for all $t \in \mathbb{R}$ and $x \in B_R(x_*(t))$, can be immediately seen by considering

$$\frac{\partial}{\partial t}(H_f(X,t)\cdot\xi) = \dot{A}(t)\xi N.$$

Thus it remains to show that the used approximation operator $G(X,t)$ of $DF(X,t)^{-1}$ satisfies for some $\tilde{c} > 0$

$$\left\|(I - DF(X,t)G(X,t))\left(\frac{1}{h}\nabla f(X,t) + G_h(X,t)(X,t)\right)\right\| \leq \tilde{c}\|\nabla f(X,t)\|, \tag{4.75}$$

for all $t \in \mathbb{R}$ and $x \in B_R(x_*(t)t)$.
It holds for $R(X,t) := \left(\frac{1}{h}\nabla f(X,t) + G_h(X,t)(X,t)\right)$

$$\|(I - DF(X,t)G(X,t))R(X,t)\| \leq \|DF(X,t)(DF(X,t)^{-1} - G(X,t))R(X,t)\| \leq$$

$$M\|DF(X,t)^{-1}R(X,t) - G(X,t)R(X,t)\|$$

for some $M > 0$. Let $v = DF(X,t)^{-1}R(X,t)$ and $w = G(X,t)R(X,t)$. Thus $v$ is a solution of (4.70), which is of the form

$$L(X,t)\cdot\eta = Q^\top R(X,t).$$

while $w$ solves (4.71), which is of the form.

$$\tilde{L}(X,t)\cdot\eta = Q^\top R(X,t).$$

Note that $L(X_*(t),t) = \tilde{L}(X_*(t),t)$ and moreover

$$(L(X,t) - \tilde{L}(X,t))\cdot\eta =$$
$$\text{offdiag }(Q^\top(A(t)-\mu I)Q)\eta N + \mu(Q^\top Q - I)(\eta D^\top D + D\eta^\top D + DD^\top\eta).$$

Let $X = X_* + \epsilon$. By construction,

$$QD = X = X_* + \epsilon,$$

for some matrix $D \in \mathbb{R}^{n\times n-p}$, defined by

$$D_{ij} = \begin{cases} \|X^i\|, & i = j \\ 0, & i \neq j \end{cases}$$

Here, $X^i$ denotes the $i$th column of $X$. Note that if $X = X_*$, then we have that $\|X^i\| = \sqrt{N_{ii}(1-\lambda_i(t)/\mu)}$, where $\lambda_i(t)$ denotes the $i$th eigenvalue of $A(t)$.
Note that $\|L(X,t) - \tilde{L}(X,t)\| = O(\|\epsilon\|)$, since the following conditions hold:

1. offdiag $(Q^\top A(t)Q) = O(\|\epsilon\|)$

   Note that $Q = [XD \;\; Y]$ with $Y = Y_* + \epsilon'$, where $\|\epsilon'\| \leq k\|\epsilon\|$ for some $k > 0$.

   Thus,

$$\text{offdiag }(Q^\top A(t)Q) = \begin{bmatrix} \text{offdiag }(D^\top X^\top A(t)XD) & D^\top X^\top A(t)Y \\ Y^\top A(t)XD & \text{offdiag }(Y^\top A(t)Y) \end{bmatrix}.$$

   Now consider the $(1,1)$ submatrix

$$\text{offdiag }(D^\top X^\top A(t)XD) = \text{offdiag }(D^\top(X_*^\top AX_* + X_*^\top A\epsilon + \epsilon^\top AX_* + \epsilon^\top A\epsilon)D) =$$
$$\text{offdiag }(D^\top(X_*^\top A\epsilon + \epsilon^\top AX_* + \epsilon^\top A\epsilon)D).$$

   The $(2,1)$ submatrix is the transposed of the $(1,2)$ submatrix, which is given by

$$D^\top X^\top A(t)Y = D^\top X_*^\top A(t)Y_* + D^\top X_*^\top A(t)\epsilon' + D^\top\epsilon A(t)Y_* + D^\top\epsilon A(t)\epsilon' =$$
$$D^\top X_*^\top A(t)\epsilon' + D^\top\epsilon A(t)Y_* + D^\top\epsilon A(t)\epsilon'.$$

   Finally the $(2,2)$ submatrix can be rewritten as

$$\text{offdiag }(Y^\top A(t)Y) = \text{offdiag }(Y_*^\top A(t)Y_* + \epsilon'^\top A(t)Y_* + Y_*A(t)\epsilon' + \epsilon'^\top A(t)\epsilon'^\top) =$$
$$\text{offdiag }(\epsilon'^\top A(t)Y_* + Y_*A(t)\epsilon' + \epsilon'^\top A(t)\epsilon'^\top).$$

   This shows that offdiag $(Q^\top A(t)Q) = O(\|\epsilon\|)$, as $|D_{ii}|$ is bounded for $1 \leq i \leq p$.

2. $Q^\top Q - I = O(\|\epsilon\|)$

   It holds
$$Q^\top Q - I = \begin{bmatrix} D^\top X^\top X D - I_p & D^\top X^\top Y \\ Y^\top X D & Y^\top Y - I_{n-p} \end{bmatrix}.$$

   By remembering that $Y$ satisfies $\|Y - Y_*\| = O(\|\epsilon\|)$, where $Y_*$ consists of the $(n-p)$ principal eigenvectors, the claim gets obvious.

Since $L(X,t)$ is well conditioned in $X = X_*$, we can conclude that the solution $w$ of the approximative system is close to the exact solution $v$, i.e.

$$\|v - w\| \leq k_1 \|X - X_*\| \leq k_2 \nabla f(X,t),$$

for some $k_1, k_2 > 0$ which shows (4.75). Note that the right estimate follows from the assumptions on $f$, cf. Theorem 2.12.

□

### 4.3.2 Principal eigenvector tracking

As mentioned above, the principal eigenvectors of $A(t)$ are minima of

$$f_-(X,t) := -f(X,t) = -\frac{1}{2}\mathrm{tr}\,(A(t)XNX^\top) - \frac{\mu}{4}\|N - X^\top X\|^2. \tag{4.76}$$

Therefore, by replacing $\nabla f$ by $-\nabla f$ and $H_f$ by $-H_f$ in Algorithm 1 and Algorithm 2, the resulting update schemes track $p$ time-varying principal eigenvectors of $A(t_k)$ for $t_k = kh$, $h > 0$. Obviously, these modified Algorithms keep their original properties. In particular, the claims regarding the accuracy of the update rules are preserved.

### 4.3.3 Numerical results

All simulations were performed in Matlab.
In the first simulation, we checked the tracking ability of algorithm (4.65) and we used approximations $G_h$ for $\frac{\partial}{\partial t}\nabla f$ of order 2. Moreover, we set $h = 0.02$, $n = 10$, $p = 3$ and $A(t) = \Theta(t)K(t)\Theta(t)^\top$, where $K(t) = \mathrm{diag}\,(a_1, ..., a_{10})$ and

$$\Theta(t) = R^\top \begin{bmatrix} \cos(t) & \sin(t) & 0 \\ -\sin(t) & \cos(t) & 0 \\ 0 & 0 & I_8 \end{bmatrix} R.$$

Here, $R \in O(10)$ is a fixed random orthogonal matrix and $a_i := 2.5i + \sin(it)$ for $i = 1, ..., 10$. We moreover used $N = \mathrm{diag}\,(p, ..., 1)$.
Figure 4.8 shows the computed (dashed) and exact (solid) 3 minor time-varying eigenvalues. As it can be seen in the corresponding error plot (Fig 4.9), where $\|X_k - X_*(t_k)\|$ is depicted, we did not use perfect initial conditions but the computed values converged fast towards the exact solution.

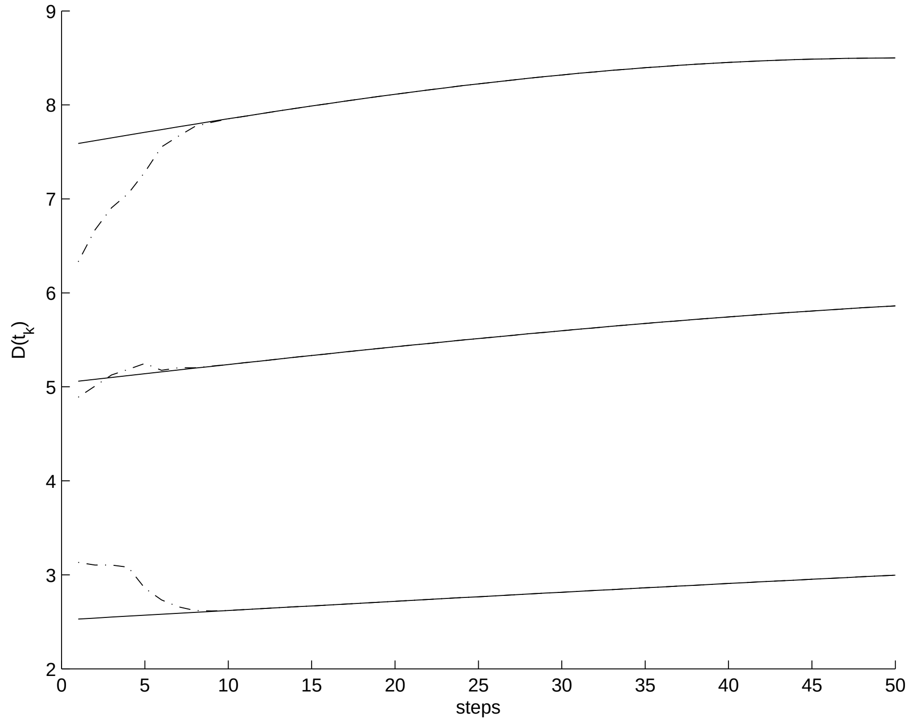


Figure 4.8: The evolution of the minor eigenvalues. Solid: exact eigenvalues, dotted: computed values of Algorithm 1.

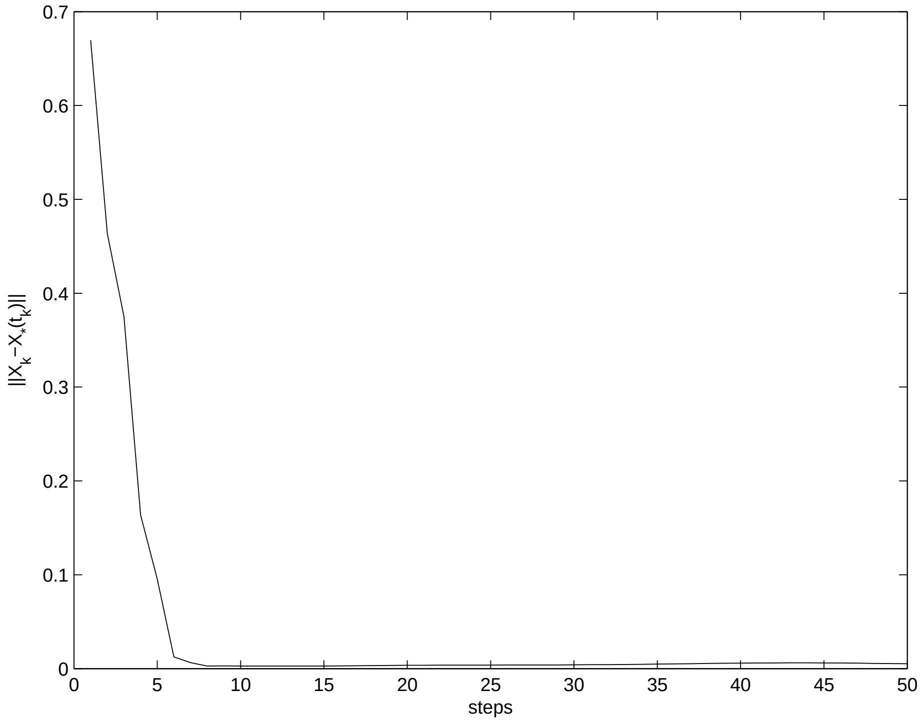


Figure 4.9: The error plot, corresponding to Figure 4.8.

| Order $m$ | Mean error Algorithm 1 | Mean error Algorithm 2 |
|---|---|---|
| 0 | $1.9 \cdot 10^{-2}$ | $1.8 \cdot 10^{-2}$ |
| 1 | $6.2 \cdot 10^{-4}$ | $5.2 \cdot 10^{-4}$ |
| 2 | $3.3 \cdot 10^{-4}$ | $2.6 \cdot 10^{-4}$ |
| 3 | $2.5 \cdot 10^{-4}$ | $2.0 \cdot 10^{-4}$ |

Table 4.5: The mean error of the two algorithms, computed for different order approximations of $\frac{\partial}{\partial t}\nabla f(X_k, t_k)$

As mentioned in the previous section, it is of much less effort to compute an approximation instead of the exact inverse of the Hessian $H_f$, which, however does not work in the general case. Thus, we replaced in the simulation $K(t)$ by $\tilde{K}(t) = \operatorname{diag}(a_1 + \sin(t), ..., a_3 + \sin(3t), a_4 + \sin(4t), ..., \ a_4 + \sin(4t)) \in \mathbb{R}^{10\times 10}$. We computed 100 steps for $h = 0.02$ and $t_{max} = 2$.
Table 4.5 shows the mean accuracy of both algorithms (4.65) and (4.73) for different choices of $G_h(x, t)$. Here the mean error is given by $\frac{1}{N}\sum_{i=1}^{N} \|X_k - X_*(t_k)\|$, where $N$ denotes the number of steps.
Hence, using approximations for $\frac{\partial}{\partial t}F(X_k, t_k)$ of order $m > 1$ significantly improves the quality of the results in both algorithms. Note further, that the second algorithm produces more precise results, although it only uses an approximatively inverted Hessian.
Next, we wanted to check the computational effort of both methods. Therefore, we made several simulations of both algorithms for different matrix dimensions $n^2$ and number of minor eigenvectors $p$. We used a second order approximation for $\frac{\partial}{\partial t}F(X_k, t_k)$ and set $K(t) = \operatorname{diag}(a_1 + \sin(t), ..., a_p + \sin(pt), \ a_{p+1} + \sin((p+1)t), ..., a_{p+1} + \sin((p+1)t) \in \mathbb{R}^{n\times n}$, $N = 100$, $h = 0.02$. We observed the computing time $t_1$ of Algorithm 1 and compared it with the elapsed time $t_2$ of Algorithm 2 for the same computation. The results in Table 4.6 show, that the second algorithm has computational advantages, which increase with the matrix size $n^2$ and the number of tracked minor eigenvectors $p$.

In our first application, we derived subspace tracking methods in Chapter 3. Since this task is closely related to the problems considered here, we wanted to compare these methods with each other. Note that the subspace tracking algorithms work under less restrictive assumptions than the algorithms derived in this chapter. Thus, we used the following setup, to assure, that Algorithm 1 and Algorithm 2 of this chapter are applicable.
We used different $n, p \in \mathbb{N}$, step size $h = 0.01$, $t_k = kh$, $k = 1, ..., 100$, and

$$A(t) = X_*(t)K(t)X_*(t)^\top \in \mathbb{R}^{n\times n},$$

for $K(t) = \operatorname{diag}(a_n, ..., a_1)$, $X_*(t) = R^\top \begin{bmatrix} \cos(t) & \sin(t) & 0 \\ -\sin(t) & \cos(t) & 0 \\ 0 & 0 & I_{n-2} \end{bmatrix} R$. Here, $R \in O(n)$ is a fixed random orthogonal matrix, $a_i := 2.5i + \sin(it)$ for $i = 1, ..., p$ and $a_i := 2.5(p+1) + \sin((p+1)t)$ for $i = p+1, ..., n$.

| $n$ | $p$ | $t_1$ (Algorithm 1) | $t_2$ (Algorithm 2) |
|---|---|---|---|
| 10 | 3 | 0.1 | 0.1 |
| 10 | 6 | 0.2 | 0.1 |
| 10 | 9 | 0.5 | 0.1 |
| 20 | 3 | 0.3 | 0.2 |
| 20 | 6 | 0.7 | 0.2 |
| 20 | 9 | 1.8 | 0.2 |
| 20 | 12 | 4.2 | 0.2 |
| 20 | 20 | 21 | 0.3 |
| 40 | 3 | 0.8 | 0.7 |
| 40 | 6 | 4.3 | 0.7 |
| 40 | 9 | 16 | 0.7 |
| 40 | 12 | 34 | 0.7 |
| 40 | 20 | 143 | 0.9 |
| 40 | 40 | 1033 | 0.9 |

Table 4.6: The computing time the two algorithms, determined for different values of $n$ and $p$

As done in Chapter 3, the task is to compute estimates $P_k$ of the principal subspace $P_*(t_k)$ of $A(t_k)$. Note that the isospectral representation $P_*$ of the principal subspace is determined by the principal eigenvectors $X_*$ via $P_* = X_* X_*^\top$. To measure the accuracy of the algorithm's output, we use the formula $\frac{1}{100}\sum_{k=1}^{100} \|P_k - P_*(t_k)\|$. Table 4.7 shows the results of different test runs, where perfect initial conditions were used. It turns out, that the subspace tracking algorithm of Chapter 3 shows the best performance regarding the accuracy and it is faster than the general MCA Algorithm 1. Note that Algorithm 2 is the fastest one, but, as mentioned before, it does not work for general subspace tracking problems.

In the case that $p = n$, Algorithm 1 and Algorithm 2 perform a complete eigenvalue

| | MCA Algorithm 1 | | MCA Algorithm 2 | | Subspace Algorithm 4 | |
|---|---|---|---|---|---|---|
| p | Comp.time | Mean error | Comp.time | Mean error | Comp.time | Mean error |
| 20 | 2.5 | $4.8 \cdot 10^{-4}$ | 0.3 | $4.7 \cdot 10^{-4}$ | 0.2 | $1.0 \cdot 10^{-5}$ |
| 40 | 20.7 | $2.3 \cdot 10^{-4}$ | 0.8 | $2.4 \cdot 10^{-4}$ | 2.7 | $1.5 \cdot 10^{-5}$ |
| 80 | 140.3 | $2.0 \cdot 10^{-4}$ | 3.2 | $1.7 \cdot 10^{-4}$ | 27.0 | $1.6 \cdot 10^{-5}$ |

Table 4.7: The computing time and mean error of the minor eigenvector tracking algorithms and the 4th subspace tracking algorithm of Chapter 3 (parameterized time-varying Newton algorithm). We used $n = 10$ and different values for $p$.

| | MCA Algorithm 2 | | EVD Algorithm | |
|---|---|---|---|---|
| n | Comp.time | Mean error | Comp.time | Mean error |
| 10 | 0.10 | $5.2 \cdot 10^{-4}$ | 0.03 | $2.0 \cdot 10^{-4}$ |
| 20 | 0.26 | $8.7 \cdot 10^{-4}$ | 0.04 | $2.7 \cdot 10^{-4}$ |
| 40 | 1.0 | $6.1 \cdot 10^{-3}$ | 0.16 | $3.2 \cdot 10^{-4}$ |

Table 4.8: The computing time and accuracy of Algorithm 2 for $p = n$ vs. the EVD algorithm of Section 4.1.

decomposition. Therefore, we finally wanted to check, whether the faster Algorithm 2 can keep up with the EVD algorithm, as described in Theorem 4.2. The calculated results show, that this is not the case, as the original EVD algorithm produces more precise results (for the same matrix $A(t)$) at lower computational costs, cf. Table 4.8.

# Chapter 5

# Application III: Pose Estimation

In this section, we consider the problem of time-varying motion reconstruction, which arises in the area of computer vision. Since we are able to formulate the particular task as a time-varying optimization problem, the previously introduced tracking algorithms are applicable. At first, we need some preparatory results.
Assume that we have a sequence of image points of a rigid object, which results by either a motion of this object or equivalently, a motion of the camera.

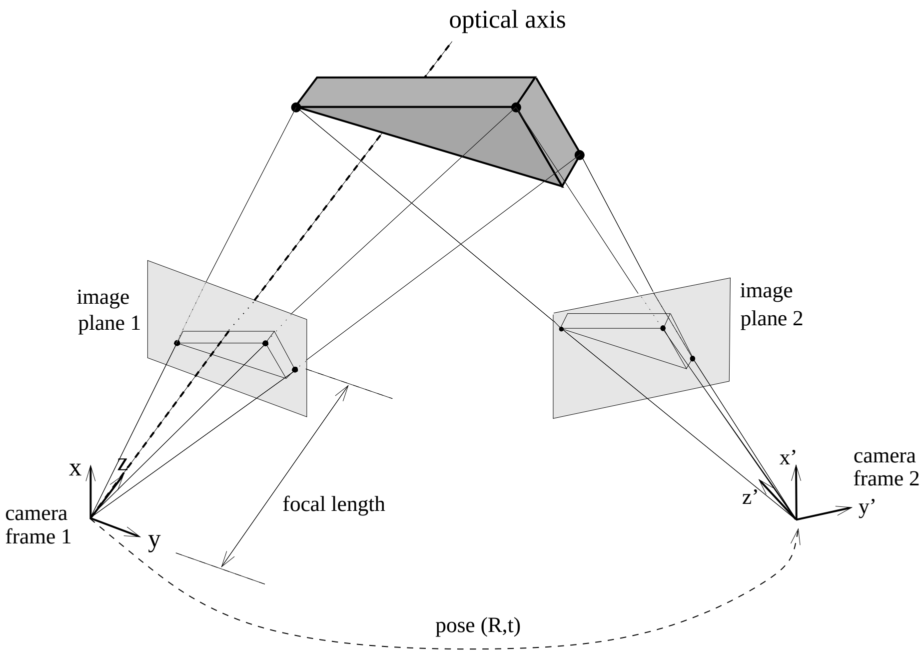


Figure 5.1: Epipolar geometry.

Thus we have for each time $t \in \mathbb{R}$ two sets of $N$ camera image points $(x'_1, ..., x'_N)$ and $(x_1(t), ..., x_N(t))$, and we assume that the correspondence between image points $x'_i \leftrightarrow x_i(t)$, $1 \leq i \leq N$, is known. The goal is to reconstruct the 6 Euclidean motion parameters of $\tau(t)$ (translation) and of $\Theta(t)$ (rotation) by evaluating the two sets of image points. As corresponding points of the same scene are related by epipolar geometry as depicted in Figure 5.1, we have for $i = 1, ..., N$ the fundamental relation (Longuet-Higgins):

$$x_i(t)^\top \Omega(t)\Theta(t)x_i' = 0, \tag{5.1}$$

where

$$\Omega(t) := \begin{bmatrix} 0 & -\tau_3(t) & \tau_2(t) \\ \tau_3(t) & 0 & -\tau_1(t) \\ -\tau_2(t) & \tau_1(t) & 0 \end{bmatrix}, \ \Theta(t) \in \mathrm{SO}(3).$$

This formula shows, that one can obtain the direction of $\tau(t) := (\tau_1(t), \tau_2(t), \tau_3(t))^\top$ but not its length by this approach: if $\hat{\Omega}\hat{\Theta}$ satisfies (5.1) for some $t \in \mathbb{R}$, then $c\hat{\Omega}\hat{\Theta}$ is as well a solution for any $c \in \mathbb{R}$.
Due to relation (5.1) we define essential matrices, which can be also considered for dimensions $n \neq 3$.

**Definition 5.1.** *An essential matrix is of the form*

$$E = \Omega\Theta,$$

*where $\Omega$ is $n \times n$ skew symmetric and $\Theta$ is $n \times n$ orthogonal.*

These essential matrices can be characterized as follows.

**Proposition 5.1.** *A matrix $E \in \mathbb{R}^{n \times n}$ admits a factorization*

$$E = \Omega\Theta, \quad \Omega \in \mathfrak{so}(\mathrm{n}),\ \Theta \in \mathrm{O}(\mathrm{n})$$

*if and only if the nonzero singular values of $E$ have even multiplicities.*
*If zero is a singular value of $E$, $\Theta$ can be chosen from* SO(3)*.*

*Proof.* "$\Rightarrow$" Let $E = \Omega\Theta$, where $\Theta \in O(n)$ and $\Omega$ is skew symmetric. Thus, the nonzero eigenvalues of $\Omega$ are all purely imaginary and come in pairs $\pm\lambda_i$ for $i = 1, ..., k$, $k \leq n/2$. This shows, that the SVD of $\Omega$ is given as

$$\Omega = U\Sigma V^\top, \tag{5.2}$$

for some $U, V \in O(n)$ and $\Sigma = \operatorname{diag}(\sigma_1, \sigma_1, \sigma_2, \sigma_2, ..., \sigma_k, \sigma_k, \underbrace{0, ..., 0}_{n-2k}) \in \mathbb{R}^{n \times n}$, where $\sigma_i = |\lambda_i|$ for $i = 1, ..., k$. Thus,

$$E = U\Sigma(V^\top\Theta)$$

is a singular value decomposition of $E$, which shows the claimed property of the singular values of $E$.
"$\Leftarrow$" Let

$$E = U\Sigma V^\top$$

for some $U, V \in O(n)$ and $\Sigma = \operatorname{diag}(\sigma_1, \sigma_1, \sigma_2, \sigma_2, ..., \sigma_k, \sigma_k, \underbrace{0, ..., 0}_{n-2k}) \in \mathbb{R}^{n \times n}$.

Define $R := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ and $B_i := \begin{bmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{bmatrix}$ for $i = 1, ..., k$. Then, $B_i R = \begin{bmatrix} \sigma_i & 0 \\ 0 & \sigma_i \end{bmatrix}$ and we get $\Sigma = B\hat{R}$, where $\hat{R} = \operatorname{diag}(\underbrace{R, ..., R}_{k}, \underbrace{1, ..., 1}_{n-2k})$ and $B = \operatorname{diag}(B_1, ..., B_k, \underbrace{0, ..., 0}_{n-2k})$.

Thus,

$$E = UBU^\top U\hat{R}V^\top,$$

where $UBU^\top$ is skew symmetric and $U\hat{R}V^\top \in O(n)$. Moreover, if $n > 2k$ then 0 is a singular value of $E$ and we can choose $U, V$ such that $U, V \in \mathrm{SO}(n)$, which implies $\Theta \in \mathrm{SO}(n)$, as $\det \hat{R} = 1$.

□

Since we are interested into smooth decompositions of time-varying essential matrices $E(t)$, we also need the following proposition.

**Proposition 5.2.** *Let $E(t) \in \mathbb{R}^{n\times n}$ with $\operatorname{rk} E(t) = 2k$ for all $t \in \mathbb{R}$ and let the non-zero singular values of $E(t)$ have even multiplicity, s.t. they are given by $(\sigma_1(t), \sigma_1(t), \sigma_2(t), \sigma_2(t), ..., \sigma_k(t), \sigma_k(t))$. We further assume, that the non-zero singular values satisfy for $i \neq j$ either $\sigma_i(t) \equiv \sigma_j(t)$ for all $t$ or $\sigma_i(t) \neq \sigma_j(t)$ for all $t$.*
*Then for smooth $t \mapsto E(t)$, there exist smooth curves $\Omega(t) \in \mathfrak{so}(\mathrm{n})$ and $\Theta(t) \in O(n)$ such that*

$$E(t) = \Omega(t)\Theta(t).$$

*Proof.* According to Proposition 4.2, there exists a SVD of $E(t)$, i.e.

$$E(t) = U(t)\Sigma(t)V(t)^\top,$$

where $U(t), \Sigma(t), V(t)$ are smooth curves in $t$.
Let $R := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, $\hat{R} = \operatorname{diag}(\underbrace{R, ..., R}_{k}, \underbrace{1, ..., 1}_{n-2k})$ and $B(t) := \Sigma(t)\hat{R}^{-1}$. Thus, $\Omega(t) := U(t)B(t)U(t)^\top$ and $\Theta(t) := U(t)\hat{R}(t)V(t)^\top$ are smooth, as well. By construction, $E(t) = \Omega(t)\Theta(t)$, which shows the claim. □

Now we have the necessary tools to reconsider the original problem (5.1), where we want to reconstruct the motion parameters $\Omega(t) \in \mathfrak{so}(3)$ and $\Theta(t) \in \mathrm{SO}(3)$.
Recall that we can not compute the length of the translation, which motivates to define the *normalized essential manifold*:

$$\varepsilon_3 := \{\Omega\Theta |\ \Omega \in \mathfrak{so}(3), \|\Omega\| = \sqrt{2}, \Theta \in \mathrm{SO}(3)\},$$

which can be equivalently characterized as

$$\varepsilon_3 = \left\{ U \begin{bmatrix} I_2 & \\ & 0 \end{bmatrix} V^\top \middle|\ U, V \in \mathrm{SO}(3) \right\}.$$

cf. Proposition 5.1. Moreover, it has already been shown in [35], that $\varepsilon_3$ is a smooth 5-dimensional manifold, which is diffeomorphic to $\{X \in \mathfrak{so}(3) |\ \|\mathrm{X}\|^2 = 2\} \times \mathrm{SO}(3)$.

By using these concepts, we can reformulate the time-varying pose estimation problem into a minimization problem: Find $E(t) \in \mathcal{E}_3$ such that

$$\frac{1}{2}\sum_{i=1}^{N}(x_i(t)^\top E(t) x_i')^2 \tag{5.3}$$

is minimal for all $t \in \mathbb{R}$. Note that in that case where the image points $x_i(t)$ are exactly determined (called *noise free case*), the minimum of (5.3) is zero, due to relation (5.1). For fixed $t \in \mathbb{R}$, this is a quadratic optimization problem on the manifold of $3 \times 3$ matrices with fixed singular values $1, 1$ and $0$, cf. above. Geometric optimization algorithms for solving this problem have been recently proposed by Helmke et al. [32] and Ma et al. [45]. For time-varying data however, such methods cannot be directly used and different approaches are required. Therefore, tracking methods basing on the Newton flow will be used to derive different algorithms to solve the minimization problem (5.3) for time-varying image points $x_i(t)$, $i = 1, ..., N$. This is equivalent to determine $(\Omega(t), \Theta(t)) \in \{X \in \mathfrak{so}(3) |\ \|\mathrm{X}\|^2 = 2\} \times \mathrm{SO}(3)$ such that

$$\Phi(\Omega, \Theta, t) = \frac{1}{2}\sum_{i=1}^{N} \operatorname{tr}^2(A_i(t)\Omega(t)\Theta(t)) \tag{5.4}$$

is minimal for all $t \in \mathbb{R}$, where $A_i(t) := x_i' x_i(t)^\top$ for $i = 1, ..., N$. Subsequently, we derive extrinsic and intrinsic methods to get a solution of this time-varying optimization problem by employing the time-varying Newton algorithm to modifications of the cost function (5.4). At the end of this chapter, numerical results demonstrate the applicability of these approaches.

## 5.1 Working in the ambient space

At first we want to solve the time-varying minimization problem (5.4) on a manifold by using an extrinsic method, i.e. we embed the algorithm into the ambient Euclidean space. Hence, we work in $\mathfrak{so}(3) \times \mathbb{R}^{3\times 3}$ (instead of $\mathcal{E}_3$), which can be identified with $\mathbb{R}^{12}$ and we define for $\mu > 0$ the modified cost function $f : \mathfrak{so}(3) \times \mathbb{R}^{3\times 3} \times \mathbb{R} \to \mathbb{R}$ by

$$f(\Omega, \Theta, t) = \frac{1}{2}\sum_{i=1}^{N} \operatorname{tr}^2(A_i(t)\Omega\Theta) + \frac{\mu}{4}(\|\Omega\|^2 - 2)^2 + \frac{\mu}{4}\|\Theta^\top\Theta - I\|^2. \tag{5.5}$$

Note that this function differs from (5.4) by including a so-called penalty term to produce a flow towards the manifold. The following two Lemmas are necessary to justify this particular choice of $f$.

**Lemma 5.1.** *For $t \in \mathbb{R}$ and $\mu > 0$ the function $f$ has compact sublevel sets and a minimum exists.*

*Proof.* For any $c > 0$ and $t = t_0 \in \mathbb{R}$,

$$c \geq f(\Omega, \Theta, t) \geq \frac{\mu}{4}(\|\Omega\|^2 - 2)^2$$

implies

$$(\|\Omega\|^2 - 2)^2 \leq 2\sqrt{\frac{c}{\mu}}$$

and thus the sublevel sets

$$f_{\leq c} := \{(\Omega, \Theta) \in \mathfrak{so}(3) \times \mathbb{R}^{3\times 3} \mid \mathrm{f}(\Omega, \Theta, \mathrm{t}_0) \leq \mathrm{c}\}$$

are compact. The result follows. $\square$

Minima of $f$ are not exactly the minima of $\Phi$ (5.4), as we added penalty terms to the original function. However, it is easy to see, that $f(\Omega, \Theta, t) = 0$ if and only if $\Phi(\Omega, \Theta, t) = 0$ and $(\Omega, \Theta) \in \varepsilon_3$, which shows, that the global minima coincide in the noise-free case.

We now prepare to apply Theorem 2.9 to the function $f$. Hence, we need the differential of $f(\Omega, \Theta, t)$ with respect to $(\Omega, \Theta)$, which is given for fixed $t$ as

$$\begin{aligned} df(\Omega, \Theta, t) \cdot (\eta, \dot{\Theta}) = \mu(\|\Omega\|^2 - 2)\mathrm{tr}\,(\Omega^\top \eta) + \mu \mathrm{tr}\,(\Theta^\top \Theta - 1)\mathrm{tr}\,(\Theta^\top \dot{\Theta}) + \\ \sum_{i=1}^{N} \mathrm{tr}\,(A_i(t)\Omega\Theta)\mathrm{tr}\,((\Theta A_i(t))_{sk}\eta) + \sum_{i=1}^{N} \mathrm{tr}\,(A_i(t)\Omega\Theta)\mathrm{tr}\,(A_i(t)\Omega\dot{\Theta}), \end{aligned} \tag{5.6}$$

where $B_{sk} := \frac{1}{2}(B - B^\top)$.

To compute a gradient of $f$, we use the standard Riemannian metric in Euclidean space. Thus for tangent vectors $\xi_1 = (\eta_1, \dot{\Theta}_1), \xi_2 = (\eta_2, \dot{\Theta}_2)$ we set

$$\langle \xi_1, \xi_2 \rangle := \mathrm{tr}\,(\eta_1^\top \eta_2) + tr(\dot{\Theta}_1^\top \dot{\Theta}_2)$$

and the gradient of $f$ is given by

$$\nabla f(\Omega, \Theta, t) = \begin{bmatrix} \mu(\|\Omega\|^2 - 2)\Omega - \sum_{i=1}^{N} \mathrm{tr}\,(A_i(t)\Omega\Theta)(\Theta A_i(t))_{sk} \\ \mu\Theta(\Theta^\top\Theta - I) - \sum_{i=1}^{N} \mathrm{tr}\,(A_i(t)\Omega\Theta)(\Omega A_i(t)^\top) \end{bmatrix}. \tag{5.7}$$

It is a necessary condition, that (5.7) is equal zero, to have a minimum of (5.5). This enables us to reformulate the original minimizing problem into a more general zero-finding problem, as isolated minima $(\Omega_*(t), \Theta_*(t))$ of (5.5) are isolated zeros of (5.7), i.e.

$$\nabla f(\Omega_*(t), \Theta_*(t), t) = 0 \quad \forall t \in \mathbb{R}. \tag{5.8}$$

The following lemma reveals some more properties of the critical points of $f$.

**Lemma 5.2.** *For $t \in \mathbb{R}$, any critical point $(\Omega_*, \Theta_*)$ of $f$ satisfies*

*(i)* $\mu(\|\Omega_*\|^2 - 2)\|\Omega_*\|^2 = -\sum_{i=1}^{N} \mathrm{tr}^2(A_i(t)\Omega_*\Theta_*)$,

*(ii)* $\mu(\|\Theta_*^\top \Theta_*\|^2 - \|\Theta_*\|^2) = -\sum_{i=1}^N \operatorname{tr}^2(A_i(t)\Omega_*\Theta_*)$.

*In particular*

$$0 < \|\Omega_*\| \le \sqrt{2}.$$

*Note that for $N > 5$ and generic $A_i(t)$, the right side of (i) and (ii) is smaller than zero, which implies $\|\Omega_*\| \lneq \sqrt{2}$.*

*Proof.* Use the critical point condition (5.8) and obtain

(i) $\mu(\|\Omega\|^2 - 2)\Omega = \sum_{i=1}^N \operatorname{tr}(A_i(t)\Omega\Theta)(\Theta A_i(t))_{sk}$,

(ii) $\mu\Theta(\Theta^\top\Theta - I) = \sum_{i=1}^N \operatorname{tr}(A_i(t)\Omega\Theta)(\Omega^\top A_i(t)^\top)$,

The result follows by multiplication on both sides with $\Omega$ in (i) and with $\Theta^\top$ in (ii) and applying the trace function to both sides of the equation. □

For the purpose of deriving the Newton flow, we determine the derivative of $\nabla f(\Omega, \Theta, t)$. Thus

$$D_1 \nabla f(\Omega, \Theta, t) \cdot \eta =$$

$$\begin{bmatrix} 2\mu \operatorname{tr}(\Omega^\top \eta)\Omega + \mu(\|\Omega\|^2 - 2)\eta - \sum_{i=1}^N \operatorname{tr}(\Theta A_i(t)\eta)(\Theta A_i(t))_{sk} \\ -\sum_{i=1}^N \operatorname{tr}(\Theta A_i(t)\eta)(\Omega A_i(t)^\top) - \sum_{i=1}^N \operatorname{tr}(A_i(t)\Omega\Theta)(\eta A_i(t)^\top) \end{bmatrix},$$

$$D_2 \nabla f(\Omega, \Theta, t) \cdot \dot\Theta =$$

$$\begin{bmatrix} -\sum_{i=1}^N \operatorname{tr}(A_i(t)\Omega\dot\Theta)(\Theta A_i(t))_{sk} - \sum_{i=1}^N \operatorname{tr}(A_i(t)\Omega\Theta)(\dot\Theta A_i(t))_{sk} \\ \mu\dot\Theta(\Theta^\top\Theta - I) + \mu\Theta(\dot\Theta^\top\Theta + \Theta^\top\dot\Theta) - \sum_{i=1}^N \operatorname{tr}(A_i(t)\Omega\dot\Theta)(\Omega A_i(t)^\top) \end{bmatrix}.$$

Let

$$H_f(\Omega, \Theta, t) := [D_1 \nabla f(\Omega, \Theta, t) \; D_2 \nabla f(\Omega, \Theta, t)] \tag{5.9}$$

denote the Hessian of $f(\Omega, \Theta, t)$. Now we are able to formulate the tracking algorithm, as a consequence of Theorem 2.9. To preserve readability, we first give a matrix-valued form of the differential equation. A way to vectorize the occurring ODE is shown later in this section.

**Theorem 5.1.** *Let $f$ as above and let $A_i(t)$ a $C^2$-map such that for some $c > 0$, $\|A_i(t)\| \le c$, $\left\|\frac{\partial}{\partial t} A_i(t)\right\| \le c$ and $\left\|\frac{\partial^2}{\partial t^2} A_i(t)\right\| \le c$ for $i = 1, ..., N$ and $t \in \mathbb{R}$. Let $t \mapsto (\Omega_*(t), \Theta_*(t))^\top$ be a smooth isolated solution of (5.8) and let $\mathcal{M}$ be a stable bundle map. Assume further, that $H_f(\Omega_*(t), \Theta_*(t), t)$ is invertible and $\|H_f(\Omega_*(t), \Theta_*(t), t)^{-1}\|$ is uniformly bounded for all $t \in \mathbb{R}$.*
*Then the solution $(\Omega(t), \Theta(t))$ of the ODE*

$$\begin{bmatrix} \dot\Omega \\ \dot\Theta \end{bmatrix} = H_f(\Omega, \Theta, t)^{-1} \Big( \mathcal{M}(\Omega\Theta) \nabla f(\Omega, \Theta, t) - \frac{\partial}{\partial t} \nabla f(\Omega, \Theta, t) \Big) \tag{5.10}$$

*converges exponentially to $(\Omega_*(t), \Theta_*(t))$, provided that $(\Omega(0), \Theta(0))$ is sufficiently close to $(\Omega_*(0), \Theta_*(0))$.*

*Proof.* As $\nabla f$ is a polynomial in $\Omega$ and $\Theta$, the assumptions regarding the derivatives of $\nabla f$ in Theorem 2.9 are obviously satisfied. The claim follows. $\square$

*Discrete tracking algorithm*
We now give a discrete version of the continuous tracking algorithm of Theorem 5.1, where a matrix-valued dynamical system is considered. Thus, before we discretize the occurring ODE, we give an equivalent vector-valued equation by employing some vectorizing operations to transform matrices into vectors, cf. appendix. Note that we particularly use different notations $\mathrm{VEC}(\cdot)$ and $\mathrm{vec}(\cdot)$, where the first one applies to all matrices, while the latter is only defined for skew symmetric matrices $X \in \mathfrak{so}(3) \subset \mathbb{R}^{3\times 3}$ by

$$\mathrm{vec}(X) := \sqrt{2}(X_{12}, X_{13}, X_{23})^\top.$$

We further need $\widetilde{\mathrm{VEC}} : \mathbb{R}^{3\times 3} \times so(3) \to \mathbb{R}^{12}$, defined by $\widetilde{\mathrm{VEC}}(\Omega, \Theta) := \begin{bmatrix} \mathrm{vec}(\Omega) \\ \mathrm{VEC}(\Theta) \end{bmatrix}$.
By using these tools, we transform the differential equation (5.10), which is defined in matrix space $so(3) \times \mathbb{R}^{3\times 3}$ into an equivalent ODE in $\mathbb{R}^{12}$, i.e.

$$\widetilde{\mathrm{VEC}}\left(\begin{bmatrix} \dot{\Omega} \\ \dot{\Theta} \end{bmatrix}\right) = H_f^{-1}(\Omega, \Theta, t)\widetilde{\mathrm{VEC}}\left(\mathcal{M}(\Omega\Theta)\nabla f(\Omega, \Theta, t) - \frac{\partial}{\partial t}\nabla f(\Omega, \Theta, t)\right), \quad (5.11)$$

where $H_f(\Omega, \Theta, t) := \begin{bmatrix} H_1 & H_2 \\ H_3 & H_4 \end{bmatrix} \in \mathbb{R}^{12\times 12}$ is the matrix representation of the Hessian of $f$ with respect to the used vectorial representation. Thus

$$\begin{bmatrix} H_1 & H_2 \\ H_3 & H_4 \end{bmatrix} \cdot \begin{bmatrix} \mathrm{vec}(\eta) \\ \mathrm{VEC}(\dot{\Theta}) \end{bmatrix} = \begin{bmatrix} \mathrm{vec}(D_1 \nabla f(\Omega, \Theta, t) \cdot \eta) \\ \mathrm{VEC}(D_2 \nabla f(\Omega, \Theta, t) \cdot \dot{\Theta}) \end{bmatrix},$$

and the submatrices of $H_f$ are given by

$$H_1 = 2\mu \mathrm{vec}(\Omega)\mathrm{vec}^\top(\Omega) + \mu(\|\Omega\|^2 - 2)I + \sum_{i=1}^{N} \mathrm{vec}((\Theta A_i(t))_{sk})\mathrm{vec}^\top((\Theta A_i(t))_{sk}) \quad (5.12)$$

$$\begin{aligned} H_2 = &\sum_{i=1}^{N} \mathrm{vec}((\Theta A_i(t))_{sk})\mathrm{VEC}^\top(\Omega A_i(t)^\top) \\ &-\frac{1}{2}\sum_{i=1}^{N} \mathrm{tr}\,(A_i(t)\Omega\Theta)L(A_i(t)^\top \otimes I - I \otimes A_i(t)^\top \pi_T) \end{aligned} \quad (5.13)$$

$$H_3 = \sum_{i=1}^{N} \mathrm{VEC}(\Omega A_i(t)^\top)\mathrm{vec}^\top((\Theta A_i(t))_{sk}) - \sum_{i=1}^{N} \mathrm{tr}\,(A_i(t)\Omega\Theta)(A_i(t) \otimes I)L' \quad (5.14)$$

$$\begin{aligned} H_4 = &\mu(\Theta^\top\Theta - I) \otimes I + \mu\Theta^\top \otimes \Theta\pi_T + \mu(I \otimes \Theta\Theta^\top) \\ &+ \sum_{i=1}^{N} \mathrm{VEC}(\Omega A_i(t)^\top)\mathrm{VEC}^\top(\Omega A_i(t)^\top) \end{aligned} \quad (5.15)$$

Here, $L$ is a linear map such that

$$L\mathrm{VEC}(\Omega) = \mathrm{vec}(\Omega)$$

for a skew symmetric matrix $\Omega$ and $\pi_T$ maps $\mathrm{VEC}(X)$ onto $\mathrm{VEC}(X^\top)$, cf. appendix. We arrive at the following discrete tracking algorithm, which is a direct consequence of Theorem 2.10. The proposed update scheme computes approximations $(\Omega_k, \Theta_k)$ of the exact minimum $(\Omega_*(t), \Theta_*(t))$ of $f$ at discrete times $t_k = kh$ for $k \in \mathbb{N}$ and $h > 0$.

**Theorem 5.2.** *Let $f$ as above and let $A_i(t)$ a $C^2$-map such that for some $c > 0$, $\|A_i(t)\| \leq c$, $\|\frac{\partial}{\partial t}A_i(t)\| \leq c$ and $\left\|\frac{\partial^2}{\partial t^2}A_i(t)\right\| \leq c$ for $i = 1, ..., N$ and all $t \in \mathbb{R}$. Let $t \mapsto (\Omega_*(t), \Theta_*(t))^\top$ satisfy (5.8) and let $G_h(\Omega, \Theta, t)$ denote an approximation of $\frac{\partial}{\partial t}\nabla f(\Omega, \Theta, t)$ satisfying for some $R, c > 0$*

$$\left\|G_h(\Omega, \Theta, t) - \frac{\partial}{\partial t}\nabla f(\Omega, \Theta, t)\right\| \leq ch$$

*for all $(\Omega, \Theta)$ with $\mathrm{dist}((\Omega, \Theta), (\Omega_*(t), \Theta_*(t))) \leq R$, $t \in \mathbb{R}$ and $h > 0$. Assume further, that the Hessian $H_f(\Omega_*(t), \Theta_*(t), t)$ is invertible and $\|H_f(\Omega_*(t), \Theta_*(t), t)^{-1}\|$ is uniformly bounded for all $t \in \mathbb{R}$.*
*Then for $c > 0$ and sufficiently small $h > 0$, the sequence*

$$\boxed{\begin{bmatrix} \Omega_{k+1} \\ \Theta_{k+1} \end{bmatrix} = \begin{bmatrix} \Omega_k \\ \Theta_k \end{bmatrix} - \widetilde{\mathrm{VEC}}^{-1}\Big(H_f(\Omega_k, \Theta_k, t_k)^{-1}\widetilde{\mathrm{VEC}}(\nabla f(\Omega_k, \Theta_k, t_k) + hG_h(\Omega_k, \Theta_k, t_k))\Big)}$$

*satisfies for $k \in \mathbb{N}$ and $t_k = kh$*

$$\|\Omega_k - \Omega_*(t_k)\|^2 + \|\Theta_k - \Theta_*(t_k)\|^2 \leq c^2h^2,$$

*provided $(\Omega_0, \Theta_0)$ is sufficiently close to $(\Omega_*(0), \Theta_*(0))$.*

## 5.2 Intrinsic pose estimation

In this section we derive an algorithm, which works directly on the manifold in order to track the minimum of the cost function (5.4).
Recall that the normalized essential manifold $\varepsilon_3$ can be described as

$$\varepsilon_3 = S \times \mathrm{SO}(3),$$

where $S := \{\Omega |\ \Omega \in so(3), \|\Omega\| = \sqrt{2}\} \cong \mathbb{S}^2$. We therefore consider the task of minimizing the function

$$\Phi : S \times \mathrm{SO}(3) \times \mathbb{R} \to \mathbb{R},$$

defined by

$$\Phi(\Omega, \Theta, t) := \frac{1}{2}\sum_{i=1}^{N} \operatorname{tr}^2(A_i(t)\Omega\Theta) \tag{5.16}$$

where $A_i(t) := x_i' x_i(t)^\top$ for $i = 1, ..., N$. In order to employ the time-varying Newton flow for this optimization problem, we need to compute the intrinsic gradient and Hessian of $\Phi$. Let therefore $(\eta_i, \Theta\psi_i) \in T_{(\Omega_i,\Theta_i)}\mathcal{E}_3,\ i = 1, 2$ denote tangent vectors of $\mathcal{E}_3$. We define a Riemannian metric by

$$\langle(\eta_1, \Theta\psi_1), (\eta_2, \Theta\psi_2)\rangle := \operatorname{tr}(\eta_1^\top \eta_2) + \operatorname{tr}(\psi_1^\top \psi_2). \tag{5.17}$$

To derive formulas for the intrinsic gradient and Hessian of $\Phi$, we will use the same approach as Ma [45], who used Newton's method to compute the minimum of $\Phi(\Omega, \Theta, t)$ for fixed $t \in \mathbb{R}$. We extend this approach by using the time-varying Newton flow in order to derive a tracking algorithm for the time-varying minimum of $\Phi$.
At first, we need the description of an orthogonal basis $B$ of the tangent space of $\mathcal{E}_3$, which is given in the following lemma.

**Lemma 5.3.** *Let* $(\Omega, \Theta) \in S \times \mathrm{SO}(3) = \mathcal{E}_3$. *Let* mat : $\mathbb{R}^3 \to \mathfrak{so}(3)$ *denote the inverse of the* vec *operation (cf. appendix) and let* $u = \operatorname{vec}(\Omega)$, $v = (v_1, v_2, v_3)^\top$ *and* $w = (w_1, w_2, w_3)^\top$ *such that* $v^\top w = 0$, $w^\top u = 0$, $v^\top u = 0$ *and* $\|v\| = \|w\| = 1$.
*Then* $\hat{B}_1 := \{\operatorname{mat}(v), \operatorname{mat}(w)\}$ *is a orthogonal basis of the tangent space* $T_\Omega S$ *of* $S$ *and a basis of the tangent space* $T_\Theta \mathrm{SO}(3)$ *of* $SO(3)$ *is given by* $\hat{B}_2 := \{\Theta \operatorname{mat}(1, 0, 0)^\top, \Theta \operatorname{mat}(0, 1, 0)^\top, \Theta \operatorname{mat}(0, 0, 1)^\top\}$.

*Proof.* As $S \subset \mathfrak{so}(3)$ can be identified with $\mathbb{S}^2 \subset \mathbb{R}^3$, we obtain the basis $\hat{B}_1$ by considering the situation in $\mathbb{S}^2$. For $u \in \mathbb{S}^2$ holds $u^\top y = 0$ if and only if $y \in T_u\mathbb{S}^2$. Therefore, we get an orthonormal basis of the tangent space $T_u\mathbb{S}^2$ by using vectors $v, w \in \mathbb{R}^3$ satisfying $v, w \neq 0$, $v^\top w = 0$, $v^\top u = 0$ and $w^\top u = 0$. Thus, $T_u\mathbb{S}^2 = \operatorname{span}(v, w)$, which shows the structure of $\hat{B}_1$. □

**Lemma 5.4.** *Let* $(\Omega, \Theta) \in \mathcal{E}_3$, *let* $\{B_1, B_2\}$ *denote an orthogonal basis of* $T_\Omega S$ *and let* $\{B_3, B_4, B_5\}$ *denote an orthogonal basis of* $T_\Theta \mathrm{SO}(3)$ *as defined in Lemma 5.3. Then for any* $t \in \mathbb{R}$, *the intrinsic gradient of* $\Phi(\Omega, \Theta, t)$ *in terms of* $B$ *is given as*

$$\operatorname{grad}\Phi(\Omega, \Theta, t) = \sum_{i=1}^{5} b_i(t) B_i,$$

*where*

$$b_i(t) = \begin{cases} \sum_{i=1}^N \operatorname{tr}(A_i(t)\Omega\Theta)\operatorname{tr}(A_i(t)B_i\Theta) & i = 1, 2 \\ \sum_{i=1}^N \operatorname{tr}(A_i(t)\Omega\Theta)\operatorname{tr}(A_i(t)\Omega B_i) & i = 3, 4, 5 \end{cases}$$

*Proof.* We use that the gradient of $\Phi$ is uniquely determined by

$$d\Phi(\Omega, \Theta, t)(\eta, \Theta\psi) = \sum_{i=1}^{N} \operatorname{tr}(A_i(t)\Omega\Theta)\operatorname{tr}(A_i(t)\eta\Theta + A_i(t)\Omega\Theta\psi) =$$

$$\langle \operatorname{grad} \Phi(\Omega, \Theta, t), (\eta, \Theta\psi) \rangle .$$

The claim gets obvious by noting that $\eta \in \operatorname{span}(B_1, B_2)$, while $\Theta\psi \in \operatorname{span}(B_3, B_4, B_5)$. □

Recall, that for fixed $t$ the Hessian form of $\Phi : \mathcal{E}_3 \to \mathbb{R}$ at a point $(\Omega, \Theta) \in \mathcal{E}_3$ is the quadratic form

$$\mathcal{H}_\Phi(\Omega, \Theta) : T_{(\Omega,\Theta)}\mathcal{E}_3 \times T_{(\Omega,\Theta)}\mathcal{E}_3 \to \mathbb{R}$$

defined for any $(\eta, \Theta\psi) \in T_{(\Omega,\Theta)}\mathcal{E}_3$ by

$$\mathcal{H}_\Phi(\Omega, \Theta)((\eta, \Theta\psi), (\eta, \Theta\psi)) := (\Phi \circ x)''(0).$$

Here $x : I \to \mathcal{E}_3$ denotes the geodesic $x(s) =: (\Omega(s), \Theta(s))$ with $(\Omega(0), \Theta(0)) = (\Omega, \Theta)$ and velocity $x'(0) = (\Omega'(0), \Theta'(0)) = (\eta, \Theta\psi)$. It is given by

$$(\Omega(s), \Theta(s)) = \left(\Omega \cos(\sigma s) + U \sin(\sigma s), \Theta \left(I + \frac{\sin(s\vartheta)}{\vartheta}\psi + \frac{1 - \cos(s\vartheta)}{\vartheta^2}\psi^2\right)\right), \quad (5.18)$$

where $\sigma = \frac{\|\eta\|}{\sqrt{2}}$ and $U = \sqrt{2}\frac{\eta}{\|\eta\|}$, $\vartheta = \sqrt{\psi_{12}^2 + \psi_{13}^2 + \psi_{23}^2}$, cf. [45].
By using these expressions, we obtain explicit formulas of the Hessian form and Hessian operator.

**Lemma 5.5.** *Let $(\Omega, \Theta) \in \mathcal{E}_3$, let $\{B_1, B_2\}$ denote an orthogonal basis of $T_\Omega S$ and let $\{B_3, B_4, B_5\}$ denote an orthogonal basis of $T_\Theta \mathrm{SO}(3)$ as defined in Lemma 5.3.
The bilinear form, associated with the Hessian form of $\Phi$ is given as*

$$\mathcal{H}_\Phi(\Omega, \Theta, t) \cdot \big((\eta_1, \Theta\psi_1), (\eta_2, \Theta\psi_2)\big) =$$

$$\frac{1}{4}\Big(\sum_{i=1}^{N} \operatorname{tr}^2(A_i(t)((\eta_1+\eta_2)\Theta+\Omega\Theta(\psi_1+\psi_2))) - \sum_{i=1}^{N} \operatorname{tr}^2(A_i(t)((\eta_1-\eta_2)\Theta+\Omega\Theta(\psi_1-\psi_2)))+$$

$$\sum_{i=1}^{N} \operatorname{tr}(A_i(t)\Omega\Theta) \operatorname{tr}\big(A_i(t)(-\Omega(\|\eta_1+\eta_2\|^2 - \|\eta_1-\eta_2\|^2)\Theta + 4\Omega\Theta\psi_1\psi_2 + 4(\eta_1\Theta\psi_2+\eta_2\Theta\psi_1))\big)\Big)$$

*Moreover, the $5 \times 5$ matrix representation with respect to the basis $\{B_1, ..., B_5\}$ of the Hessian operator of $\Phi$ is determined by*

$$(H_\Phi(\Omega, \Theta, t))_{ij} = \mathcal{H}_\Phi(\Omega, \Theta, t) \cdot (B_i, B_j), \quad 1 \le i, j \le 5.$$

*Proof.* Let $(\Omega, \Theta) \in \mathcal{E}_3$ and let $(\eta, \Theta\psi) \in T_{(\Omega,\Theta)}\mathcal{E}_3$ denote a tangent vector. Let further $(\Omega(s), \Theta(s))$ denote a geodesic on $\mathcal{E}_3$ with $(\Omega(0), \Theta(0)) = (\Omega, \Theta)$ and velocity $(\Omega'(0), \Theta'(0)) = (\eta, \Theta\psi)$, i.e. $(\Omega(s), \Theta(s))$ is as given in (5.18).
Then we consider the quadratic form $\mathcal{H}_\Phi(\Omega, \Theta, t)$

$$\mathcal{H}_\Phi(\Omega, \Theta, t) \cdot (\eta, \Theta\psi) = \Big(\Phi(\Omega, \Theta, t) \circ (\Omega(s), \Theta(s), t)\Big)''(0) =$$

$$\sum_{i=1}^{N} \operatorname{tr}^2(A_i(t)(\Omega'(0)\Theta + \Omega\Theta'(0)))+$$

$$\operatorname{tr}(A_i(t)\Omega\Theta)\operatorname{tr}\big(A_i(t)(\Omega''(0)\Theta + \Omega\Theta''(0) + 2\Omega'(0)\Theta'(0))\big) =$$

$$\sum_{i=1}^{N} \operatorname{tr}^2(A_i(t)(\eta\Theta + \Omega\Theta\psi)) + \operatorname{tr}(A_i(t)\Omega\Theta)\operatorname{tr}\big(A_i(t)(-\Omega\|\eta\|^2\Theta + \Omega\Theta\psi^2 + 2\eta\Theta\psi)\big),$$

since $\Omega'(0) = \eta$, $\Theta'(0) = \Theta\psi$, $\Omega''(0) = -\Omega\|\eta\|^2$ and $\Theta''(0) = \Theta\psi^2$.

Furthermore, the corresponding bilinear form is given by

$$\mathcal{H}_\Phi(\Omega, \Theta, t) \cdot \big((\eta_1, \Theta\psi_1), (\eta_2, \Theta\psi_2)\big) =$$

$$\frac{1}{4}\big(\mathcal{H}_\Phi(\Omega, \Theta, t) \cdot (\eta_1 + \eta_2, \Theta(\psi_1 + \psi_2)\big) - \mathcal{H}_\Phi(\Omega, \Theta, t) \cdot (\eta_1 - \eta_2, \Theta(\psi_1 - \psi_2)\big) =$$

$$\frac{1}{4}\Big(\sum_{i=1}^{N} \operatorname{tr}^2(A_i(t)((\eta_1 + \eta_2)\Theta + \Omega\Theta(\psi_1 + \psi_2)))$$

$$+\sum_{i=1}^{N} \operatorname{tr}(A_i(t)\Omega\Theta)\operatorname{tr}\big(A_i(t)(-\Omega\|\eta_1 + \eta_2\|^2\Theta + \Omega\Theta(\psi_1 + \psi_2)^2 + 2(\eta_1 + \eta_2)\Theta(\psi_1 + \psi_2))\big)$$

$$-\sum_{i=1}^{N} \operatorname{tr}^2(A_i(t)((\eta_1 - \eta_2)\Theta + \Omega\Theta(\psi_1 - \psi_2)))$$

$$-\sum_{i=1}^{N} \operatorname{tr}(A_i(t)\Omega\Theta)\operatorname{tr}\big(A_i(t)(-\Omega\|\eta_1 - \eta_2\|^2\Theta + \Omega\Theta(\psi_1 - \psi_2)^2 + 2(\eta_1 - \eta_2)\Theta(\psi_1 - \psi_2))\big)\Big) =$$

$$\frac{1}{4}\Big(\sum_{i=1}^{N} \operatorname{tr}^2(A_i(t)((\eta_1+\eta_2)\Theta+\Omega\Theta(\psi_1+\psi_2))) - \sum_{i=1}^{N} \operatorname{tr}^2(A_i(t)((\eta_1-\eta_2)\Theta+\Omega\Theta(\psi_1-\psi_2)))+$$

$$\sum_{i=1}^{N} \operatorname{tr}(A_i(t)\Omega\Theta)\operatorname{tr}\big(A_i(t)(-\Omega(\|\eta_1+\eta_2\|^2 - \|\eta_1-\eta_2\|^2)\Theta + 4\Omega\Theta\psi_1\psi_2 + 4(\eta_1\Theta\psi_2 + \eta_2\Theta\psi_1))\big)\Big)$$

The last claim follows by noting that the Hessian operator is the linear map

$$H_\Phi(\Omega, \Theta, t) : T_{(\Omega,\Theta)}\mathcal{E}_3 \to T_{(\Omega,\Theta)}\mathcal{E}_3$$

uniquely defined by

$$\mathcal{H}_\Phi(\Omega, \Theta, t)(\xi, \eta) = \langle H_\Phi(\Omega, \Theta, t)\xi, \eta\rangle \ \ \forall \xi, \eta \in T_{(\Omega,\Theta)}\mathcal{E}_3.$$

□

*Discrete tracking algorithm*
We now give a discrete tracking algorithm, which computes approximations $(\Omega_k, \Theta_k)$ of the exact pose $(\Omega_*(t), \Theta_*(t))$ at discrete times $t_k = kh$ for $k \in \mathbb{N}$ and $h > 0$. The sequence $(\Omega_k, \Theta_k)_{k \in \mathbb{N}}$ is defined as follows:

1. For $(\Omega_k, \Theta_k) \in \varepsilon_3$ choose the basis $\{B_1, B_2\}$ of $T_{\Omega_k} S$ and the basis $\{B_3, B_4, B_5\}$ of $T_{\Theta_k}\mathrm{SO}(3)$, as described in Lemma 5.3.

2. By using Lemma 5.4, compute $b := (b_1, ..., b_5)^\top$ and $c := (c_1, ..., c_5)^\top$ such that
$$\operatorname{grad}\Phi(\Omega_k, \Theta_k, t) = \sum_{i=1}^{5} b_i B_i$$
and
$$G_h(\Omega_k, \Theta_k, t) = \sum_{i=1}^{5} c_i B_i,$$
where $G_h(\Omega_k, \Theta_k, t)$ denotes an approximation for $\frac{\partial}{\partial t}\operatorname{grad}(\Omega_k, \Theta_k, t)$.

3. Determine the matrix representation of the Hessian operator $H_\Phi(\Omega_k, \Theta_k, t) \in \mathbb{R}^{5\times 5}$ with respect to the basis $\{B_1, ..., B_5\}$, cf. Lemma 5.5.

4. Compute
$$a := -H_\Phi(\Omega_k, \Theta_k, t_k)^{-1}(b + hc),$$
and set $v := a_1 B_1 + a_2 B_2$, $w = (a_3, a_4, a_5)^\top$, $\sigma = \frac{\|v\|}{\sqrt{2}}$, $U = \frac{\sqrt{2}}{\|v\|}\mathrm{mat}(v)$, $\psi = \mathrm{mat}(w)$ and $\vartheta = \sqrt{\psi_{12}^2 + \psi_{13}^2 + \psi_{23}^2}$.

5. The new point is determined by

$$\boxed{(\Omega_{k+1}, \Theta_{k+1}) = \left(\Omega_k \cos\sigma + U \sin\sigma,\ \Theta_k \left(I + \tfrac{\sin\vartheta}{\vartheta}\psi + \tfrac{1-\cos\vartheta}{\vartheta^2}\psi^2\right)\right).} \tag{5.19}$$

**Theorem 5.3.** *Let $\Phi$ as above and let $A_i(t)$ a $C^2$-map such that for some $c > 0$, $\|A_i(t)\| \le c$, $\left\|\frac{\partial}{\partial t} A_i(t)\right\| \le c$ and $\left\|\frac{\partial^2}{\partial t^2} A_i(t)\right\| \le c$ for $i = 1, ..., N$ and all $t \in \mathbb{R}$. Let $t \mapsto (\Omega_*(t), \Theta_*(t))^\top$ be a smooth isolated zero of $\operatorname{grad}\Phi(\Omega, \Theta, t)$ and let $G_h(\Omega, \Theta, t)$ denote approximation of $\frac{\partial}{\partial t}\operatorname{grad}\Phi(\Omega, \Theta, t)$ satisfying for some $R, c > 0$*
$$\left\| G_h(\Omega, \Theta, t) - \frac{\partial}{\partial t}\operatorname{grad}\Phi(\Omega, \Theta, t)\right\| \le ch$$
*for all $(\Omega, \Theta) \in \varepsilon_3$ with $\operatorname{dist}((\Omega, \Theta), (\Omega_*(t), \Theta_*(t))) \le R$, $t \in \mathbb{R}$ and $h > 0$. Assume further, that $H_\Phi(\Omega_*(t), \Theta_*(t), t)$ is invertible and $\|H_\Phi(\Omega_*(t), \Theta_*(t), t)^{-1}\|$ is uniformly bounded for $t \in \mathbb{R}$.*

*Then for $c > 0$ and sufficiently small $h > 0$, the sequence (5.19) satisfies for $k \in \mathbb{N}$ and $t_k = kh$*

$$\operatorname{dist}\big((\Omega_k, \Theta_k), (\Omega_*(t_k), \Theta_*(t_k))\big) \leq ch,$$

*provided $(\Omega_0, \Theta_0)$ is sufficiently close to $(\Omega_*(0), \Theta_*(0))$.*

*Proof.* In order to apply Theorem 2.4, we show that $\|(\pi D)^2 \operatorname{grad} \Phi(\Omega, \Theta, t)\|$ is bounded, where $\pi = \pi_{(\Omega,\Theta)}$ is the projection operator onto the tangent space of $\mathcal{E}_3$ at $(\Omega, \Theta)$. The boundedness of the other derivatives of grad $\Phi$ is obvious then.
Note that the tangent space of $\mathcal{E}_3 = S \times \mathrm{SO}(3)$ is the direct product of the tangent space of $S$ and $SO(3)$. The tangent space of $S$ at $\Omega \in S$ consists of all skew-symmetric matrices, which are orthogonal to $\Omega$. Moreover, the elements $\psi$ of the tangent space of SO(3) at $\Theta \in \mathrm{SO}(3)$ are characterized by the condition $\Theta^\top \psi \in \mathfrak{so}(3)$.
Thus for $(\eta, \psi) \in \mathfrak{so}(3) \times \mathbb{R}^{3\times 3}$, the projection operator at $(\Omega, \Theta) \in \mathcal{E}_3$ is given by

$$\pi_{(\Omega,\Theta)}(\eta, \psi) = \left(\eta - \frac{1}{2}\operatorname{tr}(\eta^\top \Omega)\Omega, \frac{1}{2}\Theta(\Theta^\top \psi - \psi^\top \Theta)\right). \tag{5.20}$$

Hence, we can compute the gradient of $\Phi$ by using the Euclidean gradient $\nabla \Phi(\Omega, \Theta, t) \in so(3) \times \mathbb{R}^{3\times 3}$, i.e.

$$\operatorname{grad} \Phi(\Omega, \Theta, t) = \pi_{(\Omega,\Theta)} \nabla \Phi(\Omega, \Theta, t).$$

Note that $\nabla \Phi(\Omega, \Theta, t)$ is determined for $(\eta, \psi) \in \mathfrak{so}(3) \times \mathbb{R}^{3\times 3}$ by

$$\langle \nabla \Phi(\Omega, \Theta, t), (\eta, \psi)\rangle = d\Phi(\Omega, \Theta, t)(\eta, \psi),$$

which is equivalent to

$$\langle \nabla \Phi(\Omega, \Theta, t), (\eta, \psi)\rangle = \sum_{i=1}^{N} \operatorname{tr}(A_i(t)\Omega\Theta) \operatorname{tr}((\Theta A_i(t))_{sk}\eta) + \sum_{i=1}^{N} \operatorname{tr}(A_i(t)\Omega\Theta) \operatorname{tr}(A_i(t)\Omega\psi).$$

This shows that

$$\nabla \Phi(\Omega, \Theta, t) = \begin{bmatrix} -\sum_{i=1}^{N} \operatorname{tr}(A_i(t)\Omega\Theta)(\Theta A_i(t))_{sk} \\ -\sum_{i=1}^{N} \operatorname{tr}(A_i(t)\Omega\Theta)\Omega A_i(t)^\top \end{bmatrix}.$$

Therefore, the intrinsic gradient is given by

$$\operatorname{grad} \Phi(\Omega, \Theta, t) = \begin{bmatrix} -\sum_{i=1}^{N} \operatorname{tr}(A_i(t)\Omega\Theta)\big((\Theta A_i(t))_{sk} + \frac{1}{2}\operatorname{tr}((\Theta A_i(t))_{sk}\Omega)\Omega\big) \\ -\sum_{i=1}^{N} \operatorname{tr}(A_i(t)\Omega\Theta)\Theta(\Theta^\top \Omega A_i(t)^\top)_{sk} \end{bmatrix}.$$

A simple calculation shows that $D\operatorname{grad}(\Omega, \Theta, t)(\psi, \eta)$ is a polynomial in $\Omega, \Theta, \psi, \eta$, where coefficients of the form $\operatorname{tr}(p(\Omega, \Theta, \psi, \eta))$ occur. Here, $p(\Omega, \Theta, \psi, \eta)$ is a polynomial in $\Omega, \Theta, \psi, \eta$.
Equation (5.20) implies that $\pi_{(\Omega,\Theta)} D\operatorname{grad}(\Omega, \Theta, t)(\psi, \eta)$ is also a polynomial in $\Omega, \Theta, \psi$, $\eta$, where coefficients of the form $\operatorname{tr}(p(\Omega, \Theta, \psi, \eta))$ occur with polynomials $p(\Omega, \Theta, \psi, \eta)$.
These reflections show that $(\pi_{(\Omega,\Theta)} D)^2 \operatorname{grad} \Phi(\Omega, \Theta, t) \cdot \big((\psi_1, \eta_1), (\psi_2, \eta_2)\big)$ is a polynomial in $\Omega, \Theta, \psi_1, \eta_1, \psi_2, \eta_2$, where coefficients of the form $\operatorname{tr}(p(\Omega, \Theta, \psi_1, \eta_1, \psi_2, \eta_2))$ occur with polynomials $p(\Omega, \Theta, \psi_1, \eta_1, \psi_2, \eta_2)$.

Hence,

$$\|(\pi_{(\Omega,\Theta)}D)^2 \mathrm{grad}\,\Phi(\Omega,\Theta,t)\cdot\big((\psi_1,\eta_1),(\psi_2,\eta_2)\big)\|$$

is bounded for $(\Omega,\Theta)\in\varepsilon_3$ and $\|\psi_1\|=\|\eta_1\|=\|\psi_2\|=\|\eta_2\|=1$, $t\in\mathbb{R}$.

□

## 5.3 Further approaches for time-varying pose estimation

Here we develop alternative methods to derive update schemes to solve the time-varying pose estimation problem. As they are closely related to the previously introduced methods, we restrict to a short description.

### 5.3.1 Partially intrinsic method

The original optimization problem is defined on the normalized essential manifold $\varepsilon_3$, which can be considered as the direct product of $S\times \mathrm{SO}(3)$, where $S := \{\Omega|\ \Omega\in\mathfrak{so}(3), \|\Omega\|=\sqrt{2}\}\cong\mathbb{S}^2$. Hence, $\varepsilon_3$ is a Riemannian submanifold of $M := \mathfrak{so}(3)\times\mathrm{SO}(3)$ and we can embed the original minimization problem (5.4) into $M$ by including penalty terms into the cost function.
Thus the optimization task turns into finding $\Omega(t)\in\mathfrak{so}(3)$ and $\Theta(t)\in\mathrm{SO}(3)$ for given matrices $A_i(t)$, $i=1,...,N$, such that for $\mu>0$

$$\Phi(\Theta,\Omega,t) := \frac{\mu}{4}(\|\Omega\|^2-2)^2+\frac{1}{2}\sum_{i=1}^{N}\mathrm{tr}^2(A_i(t)\Omega\Theta) \tag{5.21}$$

is minimal for all $t\in\mathbb{R}$.
Obviously, $\dim M=6$ and the tangent vectors at $(\Omega,\Theta)\in M$ are $(\eta,\Theta\psi)$ for $\eta,\psi\in\mathfrak{so}(3)$.
In order to minimize $\Phi$, we need to compute an intrinsic gradient. Let therefore $(\eta_i,\Theta\psi_i)\in T_{(\Omega_i\Theta_i)}M$, $i=1,2$ and define a Riemannian metric by

$$\langle(\eta_1,\Theta\psi_1),(\eta_2,\Theta\psi_2)\rangle := \mathrm{tr}\,(\eta_1^\top\eta_2)+\mathrm{tr}\,(\psi_1^\top\Theta^\top\Theta\psi_2)=\mathrm{tr}\,(\eta_1^\top\eta_2)+\mathrm{tr}\,(\psi_1^\top\psi_2). \tag{5.22}$$

The derivative of $\Phi(\Omega,\Theta,t)$ with respect to $(\Omega,\Theta)$ is given as

$$D\Phi(\Omega,\Theta,t)(\eta,\Theta\psi)=$$

$$\mu(\|\Omega\|^2-2)\mathrm{tr}\,(\Omega^\top\eta)+\sum_{i=1}^{N}\mathrm{tr}\,(A_i(t)\Omega\Theta)\mathrm{tr}\,(A_i(t)\eta\Theta+A_i(t)\Omega\Theta\psi)=$$

$$\mu(\|\Omega\|^2-2)\mathrm{tr}\,(\Omega^\top\eta)+\sum_{i=1}^{N}\mathrm{tr}\,(A_i(t)\Omega\Theta)\mathrm{tr}\,\big((\Theta A_i(t))_{sk}\eta+(A_i(t)\Omega\Theta)_{sk}\Theta^\top\Theta\psi\big),$$

where we used, that $\Theta^\top\Theta\psi\in\mathfrak{so}(3)$ and $B_{sk} := \frac{1}{2}(B-B^\top)$.

**Lemma 5.6.** *For any $t \in \mathbb{R}$, the intrinsic gradient of $\Phi(\Omega, \Theta, t)$ on $M$ is given as*

$$\operatorname{grad}\Phi(\Omega,\Theta,t) = \begin{bmatrix} \mu(\|\Omega\|^2-2)\Omega - \sum_{i=1}^N \operatorname{tr}(A_i(t)\Omega\Theta)(\Theta A_i(t))_{sk} \\ -\Theta\sum_{i=1}^N \operatorname{tr}(A_i(t)\Omega\Theta)(A_i(t)\Omega\Theta)_{sk} \end{bmatrix}. \tag{5.23}$$

*Proof.* Let for $t \in \mathbb{R}$, $\operatorname{grad}\Phi(\Omega,\Theta,t) := \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}$. Thus,

$$\operatorname{tr}(G_1^\top \eta) + \operatorname{tr}(G_2^\top \Theta\psi) =$$

$$\operatorname{tr}\left(\mu(\|\Omega\|^2-2)\Omega^\top\eta + \sum_{i=1}^N \operatorname{tr}(A_i(t)\Omega\Theta)(\Theta A_i(t))_{sk}\eta\right) +$$

$$\operatorname{tr}\left(\sum_{i=1}^N \operatorname{tr}(A_i(t)\Omega\Theta)(A_i(t)\Omega\Theta)_{sk}\Theta^\top\Theta\psi\right)$$

$$= \mu(\|\Omega\|^2-2)\operatorname{tr}(\Omega^\top\eta) + \sum_{i=1}^N \operatorname{tr}(A_i(t)\Omega\Theta)\operatorname{tr}(\Theta A_i(t)\eta) +$$

$$\sum_{i=1}^N \operatorname{tr}(A_i(t)\Omega\Theta)\operatorname{tr}(A_i(t)\Omega\Theta\Theta^\top\Theta\psi)$$

$$= \mu(\|\Omega\|^2-2)\operatorname{tr}(\Omega^\top\eta) + \sum_{i=1}^N \operatorname{tr}(A_i(t)\Omega\Theta)\operatorname{tr}(\Theta A_i(t)\eta) + \sum_{i=1}^N \operatorname{tr}(A_i(t)\Omega\Theta)\operatorname{tr}(A_i(t)\Omega\Theta\psi)$$

$$= d\Phi(\Omega,\Theta,t)(\eta,\Theta\psi).$$

□

To derive a Newton flow, we furthermore need to determine the Hessian operator which is given by the following formula:

$$H_\Phi(\Omega,\Theta,t) = \pi_{T_{(\Omega,\Theta)}M} D\operatorname{grad}\Phi(\Omega,\Theta,t), \tag{5.24}$$

where $\pi_{T_{(\Omega,\Theta)}M}$ denotes the projection operator onto the tangent space $T_{(\Omega,\Theta)}M$. We compute at first the derivative of $\operatorname{grad}\Phi$. Let therefore

$$D\operatorname{grad}\Phi(\Omega,\Theta,t)(\eta,\Theta\psi) =: \begin{bmatrix} A_1(\Omega,\Theta,t)\eta + A_2(\Omega,\Theta,t)\Theta\psi \\ B_1(\Omega,\Theta,t)\eta + B_2(\Omega,\Theta,t)\Theta\psi \end{bmatrix}.$$

Hence, the components of $D\operatorname{grad}\Phi$ are as follows:

$$A_1(\Omega,\Theta,t)\eta = 2\mu\operatorname{tr}(\Omega^\top\eta)\Omega + \eta\mu(\|\Omega\|^2-2) - \sum_{i=1}^N \operatorname{tr}((\Theta A_i(t))_{sk}\eta)(\Theta A_i(t))_{sk} \tag{5.25}$$

$$A_2(\Omega,\Theta,t)\Theta\psi = -\sum_{i=1}^{N} \operatorname{tr}\left(A_i(t)\Omega\Theta\psi\right)(\Theta A_i(t))_{sk} - \sum_{i=1}^{N} \operatorname{tr}\left(A_i(t)\Omega\Theta\right)(\Theta\psi A_i(t))_{sk} \quad (5.26)$$

$$B_1(\Omega,\Theta,t)\eta = \quad (5.27)$$

$$-\Theta\sum_{i=1}^{N} \operatorname{tr}\left((\Theta A_i(t))_{sk}\eta\right)(A_i(t)\Omega\Theta)_{sk} - \Theta\sum_{i=1}^{N} \operatorname{tr}\left(A_i(t)\Omega\Theta\right)(A_i(t)\eta\Theta)_{sk}$$

$$B_2(\Omega,\Theta,t)\Theta\psi = -\Theta\psi\sum_{i=1}^{N} \operatorname{tr}\left(A_i(t)\Omega\Theta\right)(A_i(t)\Omega\Theta)_{sk} - \quad (5.28)$$

$$\Theta\sum_{i=1}^{N} \operatorname{tr}\left(A_i(t)\Omega\Theta\psi\right)(A_i(t)\Omega\Theta)_{sk} - \Theta\sum_{i=1}^{N} \operatorname{tr}\left(A_i(t)\Omega\Theta\right)(A_i(t)\Omega\Theta\psi)_{sk}$$

Using these formulas, we obtain the following equation for the Hessian operator

$$H_\Phi(\Omega,\Theta,t)\cdot(\eta,\Theta\psi) = \begin{bmatrix} P_1(A_1(\Omega,\Theta,t)\eta + A_2(\Omega,\Theta,t)\Theta\psi) \\ P_2(B_1(\Omega,\Theta,t)\eta + B_2(\Omega,\Theta,t)\Theta\psi) \end{bmatrix},$$

where $P_1$ and $P_2$ are projection operators such that $H_\Phi(\Omega,\Theta,t)\cdot(\eta,\Theta\psi) \in T_{(\Omega,\Theta)}M$. Thus,

$$P_1(A_1(\Omega,\Theta,t)\eta + A_2(\Omega,\Theta,t)\Theta\psi) \in \mathfrak{so}(3),$$
$$P_2(B_1(\Omega,\Theta,t)\eta + B_2(\Omega,\Theta,t)\Theta\psi) \in T_\Theta \mathrm{SO}(3).$$

Since $A_1(\Omega,\Theta,t)\eta + A_2(\Omega,\Theta,t)\Theta\psi \in so(3)$, we have

$$P_1(A_1(\Omega,\Theta,t)\eta + A_2(\Omega,\Theta,t)\Theta\psi) = A_1(\Omega,\Theta,t)\eta + A_2(\Omega,\Theta,t)\Theta\psi.$$

Since $P_2$ is given by

$$P_2 : X \mapsto \Theta(\Theta^\top X)_{sk},$$

the Hessian operator turns into

$$H_\Phi(\Omega,\Theta,t)\cdot(\eta,\Theta\psi) = \begin{bmatrix} A_1(\Omega,\Theta,t)\eta + A_2(\Omega,\Theta,t)\Theta\psi \\ \Theta(\Theta^\top B_1(\Omega,\Theta,t)\eta + \Theta^\top B_2(\Omega,\Theta,t)\Theta\psi)_{sk} \end{bmatrix} =$$

$$\begin{bmatrix} A_1(\Omega,\Theta,t)\eta + A_2(\Omega,\Theta,t)\Theta\psi \\ B_1(\Omega,\Theta,t)\eta + \tilde{B}_2(\Omega,\Theta,t)\Theta\psi \end{bmatrix},$$

where

$$\tilde{B}_2(\Omega,\Theta,t)\Theta\psi = -\Theta\sum_{i=1}^{N} \operatorname{tr}\left(A_i(t)\Omega\Theta\right)(\psi(A_i(t)\Omega\Theta)_{sk})_{sk} - \quad (5.29)$$

$$\Theta\sum_{i=1}^{N} \operatorname{tr}\left(A_i(t)\Omega\Theta\psi\right)(A_i(t)\Omega\Theta)_{sk} - \Theta\sum_{i=1}^{N} \operatorname{tr}\left(A_i(t)\Omega\Theta\right)(A_i(t)\Omega\Theta\psi)_{sk}$$

By vectorizing these magnitudes (cf. appendix), we could derive a $12 \times 12$ matrix representation $\hat{H}$ of the Hessian operator. Having found a matrix representation of the Hessian, we can use the time-varying Newton flow to obtain a dynamical system to track the minimum of $\Phi$. This leads to a linear equation

$$\hat{H}_\Phi(\Omega, \Theta, t) \begin{pmatrix} \mathrm{vec}(\eta) \\ \mathrm{VEC}(\Theta\psi) \end{pmatrix} = \begin{pmatrix} \mathrm{vec}(R_1) \\ \mathrm{VEC}(R_2) \end{pmatrix}, \tag{5.30}$$

where $R_1, R_2$ are the first and second component of $\mathcal{M}(\Omega\Theta)\mathrm{grad}\,\Phi(\Omega, \Theta, t) + \frac{\partial}{\partial t}\mathrm{grad}\,\Phi$ $(\Omega, \Theta, t)$. As this linear equation is restricted to the 6 dimensional tangent space of $M$, we can transform it into a simpler form. We have

$$\hat{H}_\Phi(\Omega, \Theta, t) \begin{bmatrix} I & \\ & I \otimes \Theta \end{bmatrix} \begin{pmatrix} \mathrm{vec}(\eta) \\ \mathrm{VEC}(\Theta^\top\Theta\psi) \end{pmatrix} = \begin{bmatrix} I & \\ & I \otimes \Theta \end{bmatrix} \begin{pmatrix} \mathrm{vec}(R_1) \\ \mathrm{VEC}(\Theta^\top R_2) \end{pmatrix},$$

which is equivalent to

$$\hat{H}_\Phi(\Omega, \Theta, t) \begin{bmatrix} I & \\ & (I \otimes \Theta)L^\top \end{bmatrix} \begin{pmatrix} \mathrm{vec}(\eta) \\ \mathrm{vec}(\psi) \end{pmatrix} = \begin{bmatrix} I & \\ & (I \otimes \Theta)L^\top \end{bmatrix} \begin{pmatrix} \mathrm{vec}(R_1) \\ \mathrm{vec}(\Theta^\top R_2) \end{pmatrix},$$

where $L \in \mathbb{R}^{3\times 12}$ satisfies $\mathrm{vec}(\eta) = L\mathrm{vec}(\eta)$ for $\eta \in \mathfrak{so}(3)$, cf. appendix. This implies that

$$\underbrace{\begin{bmatrix} I & \\ & L(I \otimes \Theta^\top) \end{bmatrix} \hat{H}_\Phi(\Omega, \Theta, t) \begin{bmatrix} I & \\ & (I \otimes \Theta)L^\top \end{bmatrix}}_{=:\, \bar{H}_\Phi(\Omega, \Theta, t) \in \mathbb{R}^{6\times 6}} \begin{pmatrix} \mathrm{vec}(\eta) \\ \mathrm{vec}(\psi) \end{pmatrix} = \begin{pmatrix} \mathrm{vec}(R_1) \\ \mathrm{vec}(\Theta^\top R_2) \end{pmatrix},$$

and therefore

$$\bar{H}_\Phi(\Omega, \Theta, t) \begin{pmatrix} \mathrm{vec}(\eta) \\ \mathrm{vec}(\psi) \end{pmatrix} = \begin{pmatrix} \mathrm{vec}(R_1) \\ \mathrm{vec}(\Theta^\top R_2) \end{pmatrix}.$$

We obtain a linear equation, which has a non-singular system matrix $\bar{H}_\Phi \in \mathbb{R}^{6\times 6}$. Since $\bar{H}_\Phi$ is nothing but a matrix representation of the Hessian operator with respect to a suitable basis of the tangent space of $M$, the solutions of the reduced system also solve the original system (5.30).
By using the equations (5.25), (5.26), (5.27) and (5.29), we get the matrix representation $\bar{H}_\Phi(\Omega, \Theta, t) := \begin{bmatrix} G_1 & G_2 \\ G_3 & G_4 \end{bmatrix}$ of the Hessian operator by considering

1. $G_1\mathrm{vec}(\eta) = \mathrm{vec}(A_1(\Omega, \Theta, t)\eta)$
2. $G_2\mathrm{vec}(\psi) = \mathrm{vec}(A_2(\Omega, \Theta, t)\Theta\psi)$
3. $G_3\mathrm{vec}(\eta) = \mathrm{vec}(\Theta^\top B_1(\Omega, \Theta, t)\eta)$
4. $G_4\mathrm{vec}(\psi) = \mathrm{vec}(\Theta^\top \tilde{B}_2(\Omega, \Theta, t)\Theta\psi)$

Hence, the submatrices $G_1, ..., G_4$ of $\bar{H}_\Phi$ are given by

$$G_1 = 2\mu \mathrm{vec}\Omega \mathrm{vec}^\top \Omega + \mu(\|\Omega\|^2 - 2)I + \sum_{i=1}^{N} \mathrm{vec}(\Theta A_i(t))_{sk} \mathrm{vec}^\top (\Theta A_i(t))_{sk}, \tag{5.31}$$

$$G_2 = \sum_{i=1}^{N} \mathrm{vec}(\Theta A_i(t))_{sk} \mathrm{vec}^\top (A_i \Omega\Theta)_{sk} \tag{5.32}$$
$$-\frac{1}{2}\sum_{i=1}^{N} \mathrm{tr}\,(A_i(t)\Omega\Theta) L(A_i(t)^\top \otimes \Theta + \Theta \otimes A_i(t)^\top) L^\top,$$

$$G_3 = \sum_{i=1}^{N} \mathrm{vec}(A_i(t)\Omega\Theta)_{sk} \mathrm{vec}^\top (\Theta A_i(t))_{sk} \tag{5.33}$$
$$-\frac{1}{2}\sum_{i=1}^{N} \mathrm{tr}\,(A_i(t)\Omega\Theta) L(\Theta^\top \otimes \ A_i(t) + A_i(t) \otimes \Theta^\top) L^\top,$$

$$G_4 = \sum_{i=1}^{N} \mathrm{tr}\,(A_i(t)\Omega\Theta) L((A_i(t)\Omega\Theta)_{sk} \otimes I + I \otimes (A_i(t)\Omega\Theta)_{sk}) L^\top \tag{5.34}$$
$$+ \sum_{i=1}^{N} \mathrm{vec}(A_i(t)\Omega\Theta)_{sk} \mathrm{vec}^\top (A_i(t)\Omega\Theta)_{sk}$$
$$-\frac{1}{2} L \sum_{i=1}^{N} \mathrm{tr}\,(A_i(t)\Omega\Theta)(I \otimes (A_i(t)\Omega\Theta) + (A_i(t)\Omega\Theta) \otimes I) L^\top.$$

*Discrete tracking algorithm*

We now formulate the resulting discrete tracking algorithm, which computes approximations $(\Omega_k, \Theta_k)$ of the exact pose $(\Omega_*(t), \Theta_*(t))$ at discrete times $t_k = kh$ for $k \in \mathbb{N}$ and $h > 0$.

1. For $k \in \mathbb{N}$, compute the gradient $\mathrm{grad}\,\Phi(\Omega_k, \Theta_k, t)$ and an approximation $G_h(\Omega_k, \Theta_k, t)$ of $\frac{\partial}{\partial t}\mathrm{grad}\,(\Omega_k, \Theta_k, t)$.

2. Compute the Hessian operator $\bar{H}_\Phi(\Omega_k, \Theta_k, t) \in \mathbb{R}^{6\times 6}$ as described above and set

$$a := -\bar{H}_\Phi(\Omega_k, \Theta_k, t_k)^{-1} \begin{bmatrix} \mathrm{vec}(h(G_h)_1(\Omega_k, \Theta_k, t) + (\mathrm{grad}\,\Phi)_1(\Omega_k, \Theta_k, t)) \\ \mathrm{vec}(\Theta^\top (h(G_h)_2(\Omega_k, \Theta_k, t) + (\mathrm{grad}\,\Phi)_2(\Omega_k, \Theta_k, t))) \end{bmatrix},$$

   where $(G_h)_i(\Omega_k, \Theta_k, t)$ and $(\mathrm{grad}\,\Phi)_i(\Omega_k, \Theta_k, t)$ denote the upper or lower submatrix of $G_h(\Omega_k, \Theta_k, t)$ and $\mathrm{grad}\,\Phi(\Omega_k, \Theta_k, t)$ for $i = 1$ or $i = 2$.

3. For $\eta = \mathrm{mat}(a_1, a_2, a_3)$ and $\psi = \mathrm{mat}(a_4, a_5, a_6)$ the new point is given by

$$\boxed{(\Omega_{k+1}, \Theta_{k+1}) = \left(\Omega_k + \eta, \Theta_k \left(I + \tfrac{\sin\vartheta}{\vartheta}\psi + \tfrac{1-\cos\vartheta}{\vartheta^2}\psi^2\right)\right),} \tag{5.35}$$

where $\vartheta = \sqrt{\psi_{12}^2 + \psi_{13}^2 + \psi_{23}^2}$.

**Theorem 5.4.** *Let $M = \mathfrak{so}(3) \times \mathrm{SO}(3)$ and let $\Phi : M \to \mathbb{R}$ as above. Let $A_i(t)$ be a $C^2$-map such that for some $c > 0$, $\|A_i(t)\| \leq c$, $\left\|\frac{\partial}{\partial t}A_i(t)\right\| \leq c$ and $\left\|\frac{\partial^2}{\partial t^2}A_i(t)\right\| \leq c$ for $i = 1, ..., N$, $t \in \mathbb{R}$. Let $t \mapsto (\Omega_*(t), \Theta_*(t))^\top$ be a smooth isolated zero of* $\mathrm{grad}\,\Phi(\Omega, \Theta, t)$ *and let $G_h(\Omega, \Theta, t)$ denote an approximation of $\frac{\partial}{\partial t}\mathrm{grad}\,\Phi(\Omega, \Theta, t)$ satisfying for some $R, c > 0$*

$$\left\|G_h(\Omega, \Theta, t) - \frac{\partial}{\partial t}\mathrm{grad}\,\Phi(\Omega, \Theta, t)\right\| \leq ch,$$

*for all $(\Omega, \Theta) \in M$ with $\mathrm{dist}((\Omega, \Theta), (\Omega_*(t), \Theta_*(t))) \leq R$, $t \in \mathbb{R}$ and $h > 0$. Assume further, that $H_\Phi(\Omega_*(t), \Theta_*(t), t)$ is invertible and $\|H_\Phi(\Omega_*(t), \Theta_*(t), t)^{-1}\|$ is uniformly bounded for all $t \in \mathbb{R}$.*
*Then for $c > 0$ and sufficiently small $h > 0$, the sequence (5.35) satisfies for $k \in \mathbb{N}$ and $t_k = kh$*

$$\mathrm{dist}\big((\Omega_k, \Theta_k), (\Omega_*(t_k), \Theta_*(t_k))\big) \leq ch,$$

*provided $(\Omega_0, \Theta_0)$ is sufficiently close to $(\Omega_*(0), \Theta_*(0))$.*

*Proof.* It can be easily seen, that the conditions of Theorem 2.4 are satisfied. This shows that the sequence defined by

$$(\Omega_{k+1}, \Theta_{k+1}) = \left(\Omega_k + \eta, \exp_{\Theta_k}\psi\right)$$

satisfies

$$\mathrm{dist}\big((\Omega_k, \Theta_k), (\Omega_*(t_k), \Theta_*(t_k))\big) \leq ch$$

provided $(\Omega_0, \Theta_0)$ is sufficiently close to $(\Omega_*(0), \Theta_*(0))$. Here, $\eta$ and $\psi$ are defined as in step 3. By using Rodrigues' formula for matrix exponentials in $\mathbb{R}^{3\times 3}$, the update scheme turns into (5.35), which proves the claim.

□

### 5.3.2 Parameterization method

Again we consider the task of solving the pose estimation problem by minimizing a cost function $\Phi$ on the essential manifold. Here we use the following representation of the manifold

$$\varepsilon_3 = \left\{ U \begin{bmatrix} I_2 & \\ & 0 \end{bmatrix} V^\top \,\middle|\, U, V \in \mathrm{SO}(3) \right\},$$

and we define the cost function $\Phi : \mathcal{E}_3 \times \mathbb{R} \to \mathbb{R}$ by

$$\Phi(E,t) := \frac{1}{2}\sum_{i=1}^{N} \operatorname{tr}^2(A_i(t)E).$$

According to Helmke ([32]), a family of parameterizations $(\gamma_E)_{E\in\mathcal{E}_3}$ of the normalized essential manifold is given as follows:
Let

$$R_1 : \mathbb{R}^5 \to \mathfrak{so}(3), \begin{bmatrix} y_1 \\ \vdots \\ y_5 \end{bmatrix} \mapsto \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -\frac{y_3}{\sqrt{2}} & y_2 \\ \frac{y_3}{\sqrt{2}} & 0 & -y_1 \\ -y_2 & y_1 & 0 \end{bmatrix},$$

and

$$R_2 : \mathbb{R}^5 \to \mathfrak{so}(3), \begin{bmatrix} y_1 \\ \vdots \\ y_5 \end{bmatrix} \mapsto \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & \frac{y_3}{\sqrt{2}} & y_5 \\ -\frac{y_3}{\sqrt{2}} & 0 & -y_4 \\ -y_5 & y_4 & 0 \end{bmatrix}.$$

Let further $\Sigma = \begin{bmatrix} I_2 & \\ & 0 \end{bmatrix}$ and $U, V \in \mathrm{SO}(3)$ such that $E = U\Sigma V^\top$. Then a parameterization $\gamma_E : \mathcal{N}(0) \to \mathcal{E}_3$ of $\mathcal{E}_3$ with $\gamma_E(0) = E$ is given by

$$\gamma_E(y) := Ue^{R_1(y)}\Sigma e^{-R_2(y)}V^\top. \tag{5.36}$$

Here, $\mathcal{N}(0) \subset \mathbb{R}^5$ is a sufficiently small neighborhood of the origin.
The derivative of $\gamma_E$ reads for $h = (h_1, ..., h_5)^\top \in \mathbb{R}^5$

$$D\gamma_E(y)\cdot h = U\cdot \tag{5.37}$$

$$\left( e^{R_1(y)}\frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -\frac{h_3}{\sqrt{2}} & h_2 \\ \frac{h_3}{\sqrt{2}} & 0 & -h_1 \\ -h_2 & h_1 & 0 \end{bmatrix} \Sigma e^{-R_2(y)} - \frac{1}{\sqrt{2}} e^{R_1(y)}\Sigma e^{-R_2(y)} \begin{bmatrix} 0 & \frac{h_3}{\sqrt{2}} & h_5 \\ -\frac{h_3}{\sqrt{2}} & 0 & -h_4 \\ -h_5 & h_4 & 0 \end{bmatrix} \right)$$

$$\cdot V^\top$$

and particularly,

$$D\gamma_E(0)\cdot h = \frac{1}{\sqrt{2}} U \begin{bmatrix} 0 & -\sqrt{2}h_3 & -h_5 \\ \sqrt{2}h_3 & 0 & h_4 \\ -h_2 & h_1 & 0 \end{bmatrix} V^\top. \tag{5.38}$$

**Lemma 5.7.** *There exists a $r > 0$ such that for all $E \in \mathcal{E}_3$ the parameterization $\gamma_E$ is injective on $B_r(0) \subset \mathbb{R}^5$. In particular, there exist constants $m_1, m_2, m_3 > 0$ such that*

$$\sigma_{min}(D\gamma_E(y)) \geq m_1,$$

$$\sigma_{max}(D\gamma_E(y)) \leq m_2,$$

*and*

$$\|D^2\gamma_E(y)\| \leq m_3,$$

*for all $y \in B_r(0)$.*

*Proof.* The injectivity of the parameterization is implied if the smallest singular value of $D\gamma_E(y)$ is lower bounded by $m_1 > 0$ for all $y \in B_r(0)$.
The claim regarding the largest singular value gets obvious by considering (5.37).
To bound $\|D^2\gamma_E\|$, consider

$$D^2\gamma_E(y) \cdot (h,h) = U \cdot$$

$$\left(e^{R_1(y)}H_1^2\Sigma e^{-R_2(y)} - e^{R_1(y)}H_1\Sigma e^{-R_2(y)}H_2 - e^{R_1(y)}H_1\Sigma e^{-R_2(y)}H_2 + e^{R_1(y)}\Sigma e^{-R_2(y)}H_2^2\right)V^\top$$

where $H_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & -\frac{h_3}{\sqrt{2}} & h_2 \\ \frac{h_3}{\sqrt{2}} & 0 & -h_1 \\ -h_2 & h_1 & 0 \end{bmatrix}$ and $H_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & \frac{h_3}{\sqrt{2}} & h_5 \\ -\frac{h_3}{\sqrt{2}} & 0 & -h_4 \\ -h_5 & h_4 & 0 \end{bmatrix}$. This shows the existence of $m_2 > 0$ such that $\|D^2\gamma_E(y)\| \leq m_2$ for all $y \in \mathcal{N}(0)$ and $E \in \varepsilon_3$.
Note that (5.38) shows that $\sigma_{min}(D\gamma_E(0)) \geq c$, for some $c > 0$ and all $E \in \varepsilon_3$.
We use the following equation (Taylor):

$$D\gamma_E(y) = D\gamma_E(0) + R,$$

where $\|R\| \leq m_3\|y\|$. This shows that for $r := \frac{c}{2m_3}$, we get that

$$\sigma_{min}(D\gamma_E(y)) \geq c/2$$

for all $E \in \varepsilon_3$ and $y \in B_r(0) \cap \mathcal{N}(0)$. □

The previous lemma showed that the family of parameterizations $(\gamma_E)_{E\in\mathcal{E}_3}$ is such that we can use Main Theorem 2.3 to track the minimum $E_*(t)$ of the cost function $\Phi$. In order to do so, we still need formulas for the gradient and Hessian of $(\Phi \circ \hat{\gamma}_E)(0,t)$ w.r.t. the first component, where $\hat{\gamma}_E(y,t) := (\gamma_E(y), t)$. These magnitudes also have been computed in [32] and are given for $y \in \mathbb{R}^5$ by

$$\nabla(\Phi \circ \hat{\gamma}_E)(0,t)^\top \cdot y = \sum_{i=1}^{N} \mathrm{tr}\,(A_i(t)E)\mathrm{tr}\,\left(A_i(t)U(R_1(y)\Sigma - \Sigma R_2(y))V^\top\right) \tag{5.39}$$

and

$$y^\top H_{\Phi\circ\hat{\gamma}_E}(0,t)y = \sum_{i=1}^{N} \mathrm{tr}^{\,2}\left(A_i(t)U\left(R_1(y)\Sigma - \Sigma R_2(y)\right)V^\top\right) + \tag{5.40}$$

$$\sum_{i=1}^{N} \mathrm{tr}\,\left(A_i(t)E)\mathrm{tr}\,(A_i(t)U(R_1^2(y)\Sigma + \Sigma R_2(y)^2 - 2R_1(y)\Sigma R_2(y))V^\top\right).$$

In order to get an explicit vector-valued representation of the gradient, we vectorize the above formulas. Thus we need operations $L_1, L_2$ such that for $y \in \mathbb{R}^5$ holds

$$\mathrm{VEC}(R_1(y)) = L_1 y,$$

and

$$\mathrm{VEC}(R_2(y)) = L_2 y,$$

implying that

$$L_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and

$$L_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Then the gradient in $\mathbb{R}^5$ can be computed and is given by

$$\nabla(\Phi \circ \hat{\gamma}_E)(0,t) = \tag{5.41}$$

$$\sum_{i=1}^{N} \operatorname{tr}\left(A_i(t)E\right)\left(L_1^\top \mathrm{VEC}(U^\top A_i(t)^\top V\Sigma) - L_2^\top \mathrm{VEC}(\Sigma U^\top A_i(t)^\top V)\right).$$

Moreover, the matrix representation of the Hessian in $\mathbb{R}^{5\times 5}$ reads

$$H_{\Phi\circ\hat{\gamma}_E}(0,t) = \sum_{i=1}^{N} v_i(t)^\top v_i(t) - \operatorname{tr}\left(A_i(t)E\right) L_1^\top((U^\top A_i(t)^\top V\Sigma)\otimes I)L_1 \tag{5.42}$$

$$+\sum_{i=1}^{N} \operatorname{tr}\left(A_i(t)E\right)\left(-L_2^\top((\Sigma U^\top A_i(t)^\top V)\otimes I)L_2 + 2L_1^\top((U^\top A_i(t)^\top V)\otimes \Sigma)L_2\right),$$

where $v_i(t) = \mathrm{VEC}(U^\top A_i(t)^\top V\Sigma)^\top L_1 - \mathrm{VEC}(\Sigma U^\top A_i(t)^\top V)^\top L_2$.

Now we define the sequence $(E_k)$ tracking the minimum $E_*(t)$ of $\Phi$ at discrete times $t_k = kh$ for $k \in \mathbb{N}$, $h > 0$, by

$$\boxed{E_{k+1} = \gamma_{E_k}\left(-H_{\Phi\circ\hat{\gamma}_{E_k}}(0,t_k)^{-1}(\nabla(\Phi\circ\hat{\gamma}_{E_k})(0,t_k) + hG^h_{E_k}(0,t_k))\right),} \tag{5.43}$$

where $G^h_E(0,t)$ denotes an approximation of $\frac{\partial}{\partial t}\nabla(\Phi\circ\hat{\gamma}_E)(0,t)$.

**Theorem 5.5.** *Let $\Phi$ and $(\hat{\gamma}_E)_{E\in\mathcal{E}_3}$ as above, let $A_i(t)$ be a $C^2$-map such that for some $c > 0$, $\|A_i(t)\| \leq c$, $\left\|\frac{\partial}{\partial t}A_i(t)\right\| \leq c$ and $\left\|\frac{\partial^2}{\partial t^2}A_i(t)\right\| \leq c$ for $i = 1, ..., N$, $t \in \mathbb{R}$. Let $t \mapsto E_*(t)$ be a smooth isolated minimum of $\Phi(E,t)$ and let $G_E^h(0,t)$ denote an approximation of $\frac{\partial}{\partial t}\nabla(\Phi \circ \hat{\gamma}_E)(0,t)$ satisfying for $R, c > 0$*

$$\left\|G_E^h(0,t) - \frac{\partial}{\partial t}\nabla(\Phi \circ \hat{\gamma}_E)(0,t)\right\| \leq ch$$

*for all $E \in \mathcal{B}_R(E_*(t)) := \{E \in \varepsilon_3 | \operatorname{dist}(E, E_*(t)) < R\}$, $t \in \mathbb{R}$ and $h > 0$. Assume further, that the Hessian $H_{\Phi\circ\hat{\gamma}_E}(0,t)$ is invertible and that the norm of the inverse is uniformly bounded for all $E \in \mathcal{B}_R(E_*(t))$, $t \in \mathbb{R}$.*
*Then for $c > 0$ and sufficiently small $h > 0$, the sequence (5.43) satisfies for $k \in \mathbb{N}$ and $t_k = kh$*

$$\operatorname{dist}(E_k, E_*(t_k)) \leq ch,$$

*provided $E_0$ is sufficiently close to $E_*(0)$.*

*Proof.* We want to employ Main Theorem 2.3, where we only use one parameterization instead of two (i.e. we use $\mu_E := \gamma_E$). In order to do so, we have to show for some $\hat{R}, \tilde{R} > 0$ the boundedness of

1. $\|H_{\Phi\circ\hat{\gamma}_{E_*(t)}}(0,t)\|$ for all $t \in \mathbb{R}$,
2. $\|\frac{\partial}{\partial t}\nabla(\Phi \circ \hat{\gamma}_E)(0,t)\|$ for all $E \in \mathcal{B}_{\tilde{R}}(E_*(t))$, $t \in \mathbb{R}$,
3. $\|\frac{\partial}{\partial y}H_{\Phi\circ\hat{\gamma}_E}(y,t)\|$, $\|\frac{\partial^2}{\partial t^2}\nabla(\Phi \circ \hat{\gamma}_E)(y,t)\|$, $\|\frac{\partial}{\partial t}H_{\Phi\circ\hat{\gamma}_E}(0,t)\|$ for all $E \in \mathcal{B}_{\tilde{R}}(E_*(t))$ and $y \in B_{\hat{R}}(0)$, $t \in \mathbb{R}$.

These statements can be easily seen by considering (5.40) and computing the derivatives of (5.39) and (5.40).

□

**Remark 5.1.** Note that the update scheme (5.43) can be efficiently implemented by using additional sequences $(U_k), (V_k) \in \mathrm{SO}(3)$ as described in the sequel.

1. For $E_k = U_k\Sigma V_k^\top$, compute $\nabla(\Phi \circ \hat{\gamma}_{E_k})(0,t)$, $G_{E_k}^h(\Phi \circ \hat{\gamma}_{E_k})(0,t)$ and $H_{\Phi\circ\hat{\gamma}_{E_k}}(0,t)$ by using equations (5.41) and (5.42).
2. Solve
$$H_{\Phi\circ\hat{\gamma}_{E_k}}(0,t_k) \cdot y = -\nabla(\Phi \circ \hat{\gamma}_{E_k})(0,t_k) - hG_{E_k}^h(0,t_k)$$
for $y \in \mathbb{R}^5$.
3. Determine
$$U_{k+1} := U_k e^{R_1(y)},$$
and
$$V_{k+1} := V_k e^{R_2(y)},$$

where the occurring matrix exponentials $e^{\psi}$, $\psi \in \mathfrak{so}(3)$ can be computed by using the Rodrigues formula $e^{\psi} = I + \frac{\sin\vartheta}{\vartheta}\ \psi + \frac{1-\cos\vartheta}{\vartheta^2}\ \psi^2$, for $\vartheta = \sqrt{\psi_{12}^2 + \psi_{13}^2 + \psi_{23}^2}$.

Then the new point is given by

$$E_{k+1} = U_{k+1}\Sigma V_{k+1}^{\top}.$$

## 5.4 Numerical results

We choose random values $(x_i, y_i, z_i)^{\top} \in [-1,1] \times [-1,1] \times [2,4]$ for $i = 1, ..., N$ and $t \in \mathbb{R}$ and set $u_i' = (x_i, y_i, z_i)^{\top}$. In order to compute $v_i'(t) = \Theta(t)u_i' + \tau(t)$, we moreover used

$$\Theta(t) = R \begin{bmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \tau(t) = \begin{bmatrix} 5 - 0.1 \cdot t \\ 1 + t \\ 5 + \sin t \end{bmatrix}$$

for a fixed random orthogonal matrix $R \in \mathrm{SO}(3)$. Then the camera image points are given by $u_i = u_i'/(u_i')_3$ and $v_i(t) = v_i'(t)/(v_i'(t))_3$, where $(u_i')_3$ and $(v_i'(t))_3$ denote the 3rd entry of $u_i'$ and $v_i'(t)$, respectively.
The task was to reconstruct the rotation $\Theta(t)$ and the translation $\tau(t)$ by employing the different algorithms which evaluate $A_i(t) := u_i v_i(t)^{\top}$.
At first, we checked the tracking ability of the intrinsic algorithm as defined in Theorem 5.3. We used step size $h = 1/100$, $n = 200$ steps and perfect initial conditions to evaluate $N = 20$ image points at discrete times $t = t_k := kh$ for $k = 1, ..., n$.
Figure 5.2 depicts the error plot, i.e. the norm of the differences of the exact values of $E_*(t_k) = \Omega_*(t_k)\Theta_*(t_k)$ and the computed value $E_k = \Omega_k\Theta_k$. This shows that the computed values stay close to the exact values $E_*(t_k)$, up to a small error $< 8 \cdot 10^{-5}$.
The next graph (Figure 5.3) shows the same simulation for about 10% perturbed initial conditions. Then one observes a fast convergence of the error $\|E_k - E_*(t_k)\|$ to zero where it remains for the rest of the simulation.
At next we wanted to compare the different algorithms with each other, i.e. we used the extrinsic (Theorem 5.2), intrinsic (Theorem 5.3), partially intrinsic (Theorem 5.4) and parameterization method (Theorem 5.5) and computed the accuracy of each algorithm, where the same values for $A_i(t_k)$ and exact initial values were used for each method.
The results are shown in Table 5.1, Table 5.2 and Table 5.3, where we used different numbers of points $N$ or different step sizes $h$ in each table. In the tables, the mean error is defined as

$$\frac{1}{n}\sum_{k=1}^{n} \|E_k - E_*(t_k)\|,$$

where $n$ denotes the number of steps. It turns out, that all presented methods are able to track the desired transformation. The parameterization method however, has significant computationally advantages regarding computing time, while the intrinsic algorithms work at a higher accuracy.
These simulations also confirm the expectation, that increasing the number of evaluated points $N$ or decreasing the step size $h$ improves the accuracy of each algorithm.

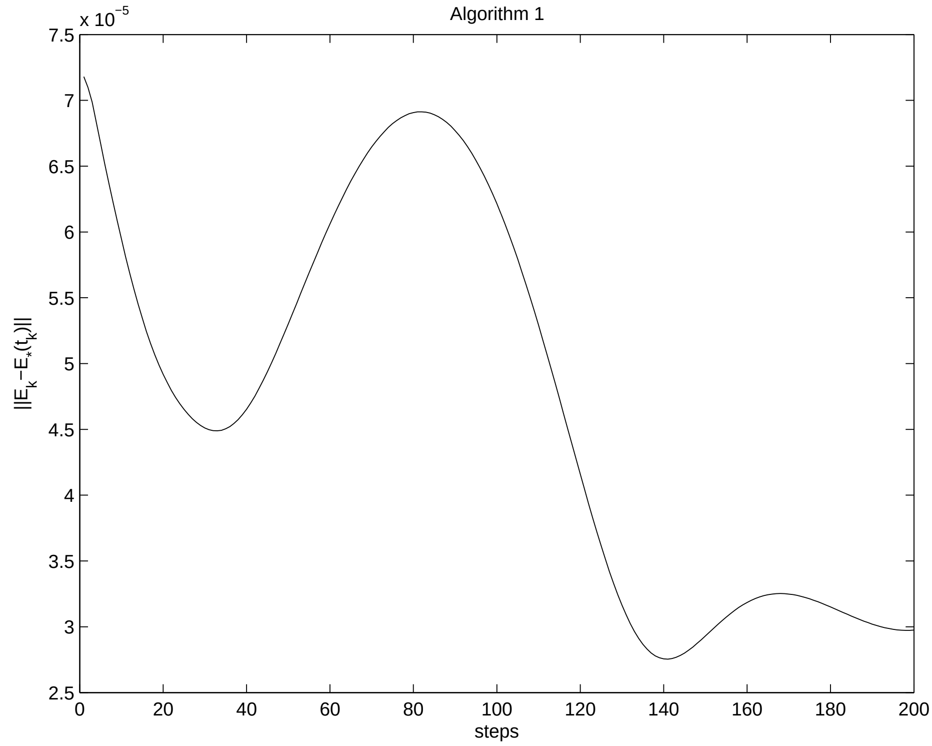


Figure 5.2: The error $\|E_k - E_*(t_k)\|$ during the algorithm, where perfect initial conditions were used.

| Method | Computing time | Mean error |
|---|---|---|
| Extrinsic | 1.0 | $1.8 \cdot 10^{-2}$ |
| Intrinsic | 6.3 | $4.3 \cdot 10^{-4}$ |
| Partially intrinsic | 1.7 | $4.1 \cdot 10^{-4}$ |
| Parameterization | 0.9 | $5.5 \cdot 10^{-4}$ |

Table 5.1: The computing time and mean error of different pose estimation methods. We used $N = 10$, $n = 100$ and $h = 0.01$.

| Method | Computing time | Mean error |
|---|---|---|
| Extrinsic | 1.9 | $7.4 \cdot 10^{-3}$ |
| Intrinsic | 12.8 | $6.7 \cdot 10^{-5}$ |
| Partially intrinsic | 3.4 | $5.5 \cdot 10^{-5}$ |
| Parameterization | 1.7 | $1.2 \cdot 10^{-4}$ |

Table 5.2: The computing time and mean error of different pose estimation methods. We used $N = 20$, $n = 100$ and $h = 0.01$.

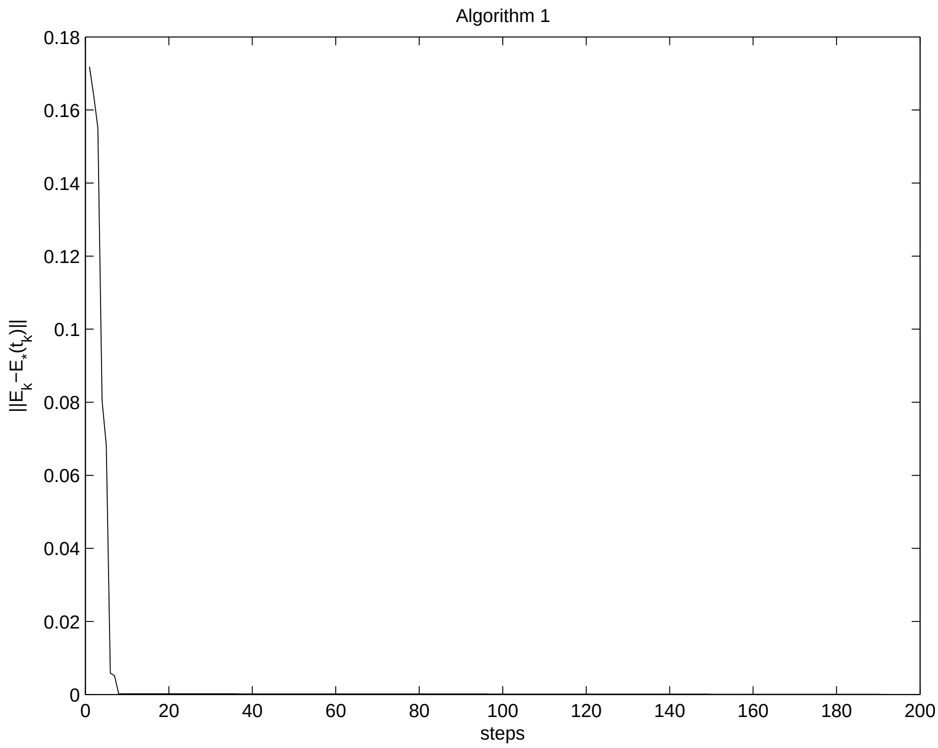


Figure 5.3: The error $\|E_k - E_*(t_k)\|$ during the algorithm. Here, perturbed initial values were used.

| Method | Computing time | Mean error |
|---|---|---|
| Extrinsic | 1.9 | $7.6 \cdot 10^{-4}$ |
| Intrinsic | 12.6 | $1.5 \cdot 10^{-5}$ |
| Partially intrinsic | 3.3 | $1.5 \cdot 10^{-5}$ |
| Parameterization | 1.7 | $1.8 \cdot 10^{-5}$ |

Table 5.3: The computing time and mean error of different pose estimation methods. We used $N = 20$, $n = 100$ and $h = 0.005$.

# Chapter 6

# Appendix

Here we give some tools which are necessary to work with vectorized matrices.
At first we introduce the VEC-Operation in $\mathbb{R}^{m\times n}$ which transforms matrices into vectors by stacking the columns of the matrix under each other, cf. Horn and Johnson [36]. Thus for $X = [X^1 \,...\, X^n] \in \mathbb{R}^{m\times n}$,

$$\mathrm{VEC}(X) := \begin{bmatrix} X^1 \\ \vdots \\ X^n \end{bmatrix} \in \mathbb{R}^{mn}.$$

To deal with skew symmetric matrices $X \in \mathfrak{so}(\mathrm{n}) \subset \mathbb{R}^{\mathrm{n}\times\mathrm{n}}$, we define a modified operation vec : $\mathfrak{so}(\mathrm{n}) \to \mathbb{R}^{\mathrm{n}\times\mathrm{n}}$, which maps skew-symmetric matrices bijectively onto $\mathbb{R}^{\frac{n(n-1)}{2}}$ by

$$\mathrm{vec}(X) := \sqrt{2} \begin{bmatrix} \tilde{X}^1 \\ \vdots \\ \tilde{X}^{n-1} \end{bmatrix} \in \mathbb{R}^{\frac{n(n-1)}{2}}.$$

Here, $\tilde{X}^i$ denotes the upper diagonal part of the $(i+1)$th column of $X$, i.e. $\tilde{X}^1 = X_{1,2}, \tilde{X}^2 = (X_{1,3}, X_{2,3})^\top, ..., X^{n-1} = (X_{1,n}, ..., X_{n-1,n})^\top$. Note that we used the $\sqrt{2}$-factor to ensure

$$\|\mathrm{VEC}(X)\| = \|\mathrm{vec}(X)\|,$$

where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^n$.
For the particular case $n = 3$, we need the inverse of the vec operation, denoted by mat : $\mathbb{R}^3 \to \mathfrak{so}(3)$. Thus for matrices

$$\Omega = \begin{bmatrix} 0 & \Omega_1 & \Omega_2 \\ -\Omega_1 & 0 & \Omega_3 \\ -\Omega_2 & -\Omega_3 & 0 \end{bmatrix} \in \mathfrak{so}(3),$$

we have

$$\Omega = \sqrt{2}\ \mathrm{mat}(\Omega_1, \Omega_2, \Omega_3)^\top.$$

In the area of computer vision, we further need to vectorize tuples of matrices $(S, A) \in \mathfrak{so}(3) \times \mathbb{R}^{3\times 3}$, which we define as follows

$$\widetilde{\mathrm{VEC}}\left(\begin{bmatrix} S \\ A \end{bmatrix}\right) := \begin{bmatrix} \mathrm{vec}(S) \\ \mathrm{VEC}(A) \end{bmatrix} \in \mathbb{R}^{12}.$$

To vectorize matrix-valued functions, we moreover use the following known formulas for matrices $X \in \mathbb{R}^{m\times n}$, $Y \in \mathbb{R}^{n\times r}$, $Z \in \mathbb{R}^{r\times s}$

$$\mathrm{VEC}(XY) = (I \otimes X)\mathrm{VEC}(Y) = (Y^\top \otimes I)\mathrm{VEC}(X), \tag{6.1}$$

$$\mathrm{VEC}(XYZ) = (Z^\top \otimes X)\mathrm{VEC}(Y), \tag{6.2}$$

where $\otimes$ denotes the Kronecker product. We further need an operation $L$ such that for any skew symmetric matrix $S \in \mathfrak{so}(\mathrm{n})$ holds

$$\mathrm{vec}(S) = L\mathrm{VEC}(S) \tag{6.3}$$

and

$$\mathrm{VEC}(S) = L^\top \mathrm{vec}(S). \tag{6.4}$$

Note, that $LL^\top = I$ and for $n = 3$,

$$L = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \end{bmatrix}.$$

Note further, that for $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$ holds

$$\mathrm{tr}\,(AB) = \mathrm{VEC}^\top(A^\top)\mathrm{VEC}(B), \tag{6.5}$$

and for $A \in \mathbb{R}^{n\times n}$, $S \in \mathfrak{so}(\mathrm{n})$

$$\mathrm{tr}\,(AS) = \mathrm{tr}\,(A_{sk}S) = \mathrm{vec}^\top(A_{sk}^\top)\mathrm{vec}(S) = -\mathrm{vec}^\top(A_{sk})\mathrm{vec}(S). \tag{6.6}$$

Here, $A_{sk} = \frac{1}{2}(A - A^\top)$ denotes the skew symmetric part of $A$.
Finally, we need a matrix $\pi \in \mathbb{R}^{mn\times mn}$ such that

$$\pi\mathrm{VEC}(Z) = \mathrm{VEC}(Z^\top),$$

for all $Z \in \mathbb{R}^{m\times n}$. Hence, $\pi$ is a permutation matrix with 1's at positions

$$\big((i-1)n + j,\ (j-1)m + i\big), \quad \text{for } 1 \le i \le m,\ 1 \le j \le n.$$

In the case $n = m = 3$, $\pi$ is given by

$$\pi = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

# Bibliography


[1] R. Abraham, J.E. Marsden, T. Ratiu, *Manifolds, Tensor Analysis, and Applications*, Springer-Verlag New York, 1988.

[2] P.-A. Absil, R. Mahony, and R. Sepulchre, Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematicae*, 80:199–220, 2004.

[3] R.L. Adler, J.-P. Dedieu, J.Y. Margulies, M. Martens, M. Shub, Newton's Method on Riemannian manifolds and a geometric Model of the human Spine, *IMA J. of Num. Analysis* 22:359-390, 2002.

[4] E.L. Allgower, K. Georg, Continuation and Path Following, *Acta Numerica*, 1992.

[5] E.L. Allgower, K. Georg, *Introduction to Numerical Continuation Methods*, SIAM, 2003.

[6] M. Baumann, U. Helmke, Singular value decomposition of time-varying matrices, *Future Generation Computer Systems* 19, 2003.

[7] M. Baumann, U. Helmke, A path-following method for tracking of time-varying essential matrices, *Proceedings of the MTNS*, CD-Rom, Leuven, 2004.

[8] M. Baumann, C. Lageman, U. Helmke, Newton-type algorithms for time-varying pose estimation, *Proceedings of the ISSNIP*, CD-Rom, Melbourne, 2004.

[9] S.J.G. Bell, N.K. Nichols, A numerical method for the computation of the analytic singular value decomposition, *Num. Analysis Rep.* 2, 1994.

[10] Ben-Israel, A., A Newton-Raphson method for the solution of systems of equations, *Journal of Mathematical Analysis and Applications* 15, 1966.

[11] H.W. Braden, The equations $A^\top X \pm X^\top A = B$, *SIAM Journal on Matrix Analysis and Applications*, 20:295-302, 1998.

[12] F.H. Branin, Widely convergent Method for Finding multiple Solutions of simultaneous nonlinear Equations, *IBM Journal of Research and Development* 16:504-521, 1972.

[13] R.W. Brockett, Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems, *Linear Algebra Appl.*, 146:79-91, 1991.

[14] A. Bunse-Gerstner, R. Byers, V. Mehrmann, N.K. Nichols, Numerical Computation of an analytical Singular Value Decomposition of a Matrix valued Function, *Num. Mathematik*, 60:1-39, 1991.

[15] P. Comon, G.H. Golub, Tracking a few extreme singular values and vectors in signal processing, *Proceedings of the IEEE*, 78(8):1327-1343, 1990.

[16] D. Davidenko, On a new method of numerically integrating a system of nonlinear equations, *Dokl. Akad. Nauk SSSR*, 88:601-604, 1953.

[17] J.-P. Dedieu, P. Priouret, Newton's method on Riemannian manifolds: covariant alpha theory, *IMA J. of Num. Analysis* 23:395-419, 2003.

[18] J.-P. Dedieu, M. Shub, Newton's Method for overdetermined Systems of Equations, *Mathematics of Computation* 69:1099-1115, 1999.

[19] R.S. Dembo, S.C. Eisenstat, T. Steihaug, Inexact Newton methods, *SIAM J. Numer. Anal.* 19:400-408, 1982.

[20] L. Dieci, T. Eirola, On smooth decompositions of matrices, *SIAM Journal on Matrix Analysis and Applications*, 20:800-819, 1999.

[21] M.P. do Carmo, *Riemannian Geometry*, Birkhäuser Boston, 1992.

[22] O. Faugeras, Three-dimensional computer vision: a geometric viewpoint, *The MIT Press*, 1993.

[23] A. Edelman, T. A. Arias, S. T. Smith, The geometry of algorithms with orthogonality constraints, *SIAM J. Matrix Anal. Appl.* 20:303-353, 1998.

[24] O.P. Ferreira, B.F. Svaiter, Kantorovich's Theorem on Newton's Method in Riemannian Manifolds, *Journal of Complexity* 18:304-329, 2002.

[25] D. Gabay, Minimizing a differentiable Function over a Differential Manifold, *Journal of Optimization Theory and Applications* 37:177-219, 1982.

[26] C.B. Garcia, F.J. Gould, Relations between several Path following Algorithms and local and global Newton Methods, *SIAM Review* 22(3):263-274, 1980.

[27] C. Geiger, C. Kanzow, *Theorie und Numerik restringierter Optimierungsaufgaben*, Springer-Verlag Berlin, 2002.

[28] N.H. Getz, *Dynamic inversion of nonlinear maps*, Ph.D. Thesis, Univ. of California at Berkeley, 1995.

[29] D. Gromoll, W. Klingenberg, W. Meyer, *Springer LNM 5: Riemannsche Geometrie im Großen*, Springer-Verlag, Berlin, 1968

[30] R. Hartley, A. Zisserman, *Multiple View Geometry*, Cambridge University Press, 2000.

[31] S. Helgason, *Differential geometry, Lie groups and symmetric spaces*, Academic Press London, 1978.

[32] U. Helmke, K. Hüper, P.-Y. Lee, J. Moore, Essential matrix estimation via alternating projections, preprint, 2003.

[33] U. Helmke, K. Hüper, J. Trumpf, Newton's Method on Grassmanian Manifolds, preprint, 2006.

[34] U.Helmke, P.S. Krishnaprasad, Algorithmic Mechanics on Lie Groups, *Preprint*, 2005.

[35] U. Helmke, J.B. Moore, *Optimization and Dynamical Systems*, Springer-Verlag London, 1994.

[36] R.A. Horn, C.R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, 1991.

[37] J. Huitfeldt, A. Ruhe, A new Algorithm for numerical Path following applied to an Example from hydrodynamical Flow, *SIAM J. Sci. Stat. Comput.* 11(6):1181-1192, 1990.

[38] K. Hüper and J. Trumpf, Newton-like methods for numerical optimization on manifolds, *Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, 2004.

[39] H.T. Jongen, P. Jonker, F.Twilt, *Nonlinear Optimization in Finite Dimensions*, Kluwer Academic Publishers Dordrecht, 2000.

[40] Z. Kang, C. Chatterje, V.P. Roychowdhury, An adaptive quasi-Newton algorithm for eigensubspace estimation, *IEEE Transactions on Sig. Proc.*, 48(12):3328-3333, 2000.

[41] T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag Berlin, 1966.

[42] A.W. Knapp, *Lie Groups beyond an Introduction*, Birkhäuser Berlin, 2002.

[43] A.J. Laub, M.T. Heath, C.C. Paige, R.C. Ward, Computation of system balancing transformations and other applications of simultaneous diagonalization algorithms, *IEEE Transactions Automatic Control* 32:115-121, 1987.

[44] E. Lundström, L. Elden, Adaptive eigenvalue computations using Newton's method on the Grassmann manifold, *SIAM J. Matrix Anal. Appl.*, 23(3):819839, 2002.

[45] Y. Ma, J. Koseca, S. Sastry, Optimization criteria and geometric algorithms for motion and structure estimation, *Intern. J. of Computer Vision* 44, 2001.

[46] R. Mahony, J.H. Manton, The geometry of the Newton method on non-compact Lie groups, *Journal of Global Optimization* 23:309-327, 2002.

[47] J.H. Manton, On the various generalizations of optimization algorithms to manifolds, *Proceedings of the MTNS*, CD-Rom, Leuven, 2004.

[48] J. H. Manton, U. Helmke, I.M.Y. Mareels, A dual purpose principal and minor component flow, *Systems and Control Letters* 54:759-769, 2005.

[49] J.-P. Laumond, *Robot Motion Planning and Control*, Lecture Notes in Control and Inf. Sciences 229, Springer-Verlag London, 1998.

[50] G. Mathev, V.U. Reddy, S. Dasgupta, Adaptive estimation of eigenspace, *IEEE Transactions on Sig. Proc.* 43(2):401-411, 1995.

[51] V. Mehrmann, W. Rath, Numerical methods for the computation of analytic singular value decompositions, Electr. Trans. Numer. Anal. 1:72-88, 1992.

[52] C. B. Moler and C. F. Van Loan, Nineteen Dubious Ways to Compute the Exponential of a Matrix, *SIAM Review* 20, 1979.

[53] J.J. Murray, C.H. Lo, Continuation Methods for Systems with linear Constraints, *Linear Circuits, Systems and Signal Processing* 73:80, 1988.

[54] S. Nicosia, A. Tornambe, P. Valigi, A solution to the generalized problem of nonlinear map inversion, *System and Control Letters* 17(5), 1991.

[55] E. Oja, H. Ogawa, J. Wangviwattana, Principal component analysis by homogenous neural networks, Part I: The weighted subspace criterion, *IEICE Transactions on Information and Systems* 3:366-375, 1992.

[56] E. Oja, H. Ogawa, J. Wangviwattana, Principal component analysis by homogenous neural networks, Part II: Analysis and extensions of the learning algorithms, *IEICE Transactions on Information and Systems* 3:376-382, 1992.

[57] A.B. Poore, A continuation approach to constraint optimization, *Analysis and Control of Nonlinear Systems* 181:186, 1988.

[58] D.J. Rabideau, Fast rank adaptive subspace tracking and applications, *IEEE Transactions on Signal Processing* 44(9), 1996.

[59] E.C. Real, D.W. Tufts, J.W. Cooley, Two algorithms for fast approximative subspace tracking, *IEEE Transactions on Signal Processing* 47(7), 1999.

[60] S.L. Richter, Continuation Methods: Theory and Applications, *IEEE Trans. Circ. and Syst.* 347:352, 1983.

[61] M. Shub. Some remarks on dynamical systems and numerical analysis. In *Dynamical systems and partial differential equations (Caracas, 1984)*, pages 69–91. Univ. Simon Bolivar, Caracas, 1986.

[62] Y. Sibuya, Some global properties of matrices of functions of one variable, *Math. Ann.* 161:67-77, 1965.

[63] S. Smith, Optimization Techniques on Riemannian Manifolds, *Fields Institute Communications* 3:113-136, 1994.

[64] J. Stoer, R. Bulirsch, *Numerische Mathematik 2*, Springer Verlag Berlin, 1990.

[65] K. Tanabe, Differential Geometric Methods in nonlinear Programming, *Applied Nonlinear Analysis (Proc. Third Internat. Conf., Univ. Texas, Arlington, Tex.)* pp.707-719, Academic Press New York-London, 1979.

[66] K. Tanabe, Continuous Newton-Raphson Method for solving an underdetermined System of nonlinear Equations, *Nonlinear Analysis, Theory, Methods & Appl.* 3(4):495-503, 1979.

[67] K. Tchon, I. Duleba, On inverting singular kinematics and geodesic trajectory generation for robotic manipulators, *Journal of Intelligent and Robotic Systems* 8, 1993.

[68] C. Udistre, *Convex Functions and Optimization Methods on Riemannian Manifolds*, Kluwer Academic Publishers Dordrecht, 1994.

[69] X. Wang, H.V. Poor, Blind multiuser detection: A subspace approach, *IEEE Transactions on Information Theory* 44(2), 1998.

[70] K. Wright, Differential equations for the analytical singular value decomposition of a matrix, *Numer. Math.* 63:283-295, 1992.

[71] L. Xu, Least mean square error recognition principle for self organizing neural nets, *Neural Networks* 6, 1993.

[72] B. Yang, Projection approximation subspace tracking, *IEEE Transactions on Signal Processing* 43(1), 1995.

[73] J.F. Yang, M. Kaveh, Adaptive eigenspace algorithms for direction or frequency estimation and tracking, *IEEE Transactions on Acoustics, Speech and Signal Processing* 36(2), 1988.

[74] S. Yoshizawa, U. Helmke, K. Starkov, Convergence analysis for principal component flows, *International Journal of applied Mathematics and Computer Science* 11, 2001.