Lie group methods *

Elena Celledoni†
November 3, 2009

1 Background material

This is section a summary of the results of the first and second chapter in [Ol] which are particularly relevant for the introduction of Lie group methods.

1.1 Manifolds

Definition 1.1. [Ol] An m-dimensional manifold \( \mathcal{M} \) is a topological space covered by a collection of open subsets \( W_\alpha \subset \mathcal{M} \) (coordinate charts) and maps \( \mathcal{X}_\alpha : W_\alpha \rightarrow V_\alpha \subset \mathbb{R}^m \) one-to-one and onto, where \( V_\alpha \) is an open, connected subset of \( \mathbb{R}^m \). \( (W_\alpha, \mathcal{X}_\alpha) \) define coordinates on \( \mathcal{M} \).

\( \mathcal{M} \) is a smooth manifold if the maps \( \mathcal{X}_{\alpha\beta} = \mathcal{X}_\beta \circ \mathcal{X}_\alpha^{-1} \), are smooth where they are defined, i.e. on \( \mathcal{X}_\alpha(W_\alpha \cap W_\beta) \) to \( \mathcal{X}_\beta(W_\alpha \cap W_\beta) \).

Example 1.2. \( \mathbb{R}^m \) is a m-dimensional manifold covered with a single chart.

Example 1.3. The unit sphere \( S^{m-1} := \{ x \in \mathbb{R}^m | \sum_{i=1}^m x_i^2 = 1 \} \) is a \( m-1 \)-dimensional manifold covered with two charts obtained by omitting the north and south poles respectively. The coordinate maps are obtained considering the stereographic projection from the north and south pole respectively.

Given two smooth manifolds \( \mathcal{M} \) and \( \mathcal{N} \) we say that \( F : \mathcal{M} \rightarrow \mathcal{N} \) is a smooth map if it is smooth in local coordinates. Introducing local coordinates on the manifolds we get \( x \in \mathcal{M}, x = (x_1, \ldots, x_m), \) and \( y \in \mathcal{N}, y = (y_1, \ldots, y_n) \). Assume \( y = F(x) \) and \( y_i = F_i(x) \ i = 1, \ldots, n; \) if \( F_i \) is smooth as a map from an open subset of \( \mathbb{R}^m \) to \( \mathbb{R} \), then \( F \) is smooth also as map between \( \mathcal{M} \) and \( \mathcal{N} \).

*Copyright E. Celledoni, NTNU, 2008, all rights reserved
†Department of Mathematical Sciences NTNU
The rank of a map $F : \mathcal{M} \to \mathcal{N}$ is the rank of the Jacobian of $F$, the map is said to be regular if its rank is constant.

A subset $\mathcal{N} \subset \mathcal{M}$ of a manifold which is a manifold in its own right is a submanifold.

**Definition 1.4. Submanifolds** An immersed submanifold $\mathcal{N}$ of a manifold $\mathcal{M}$ is a subset $\mathcal{N} \subset \mathcal{M}$ and a map $F$ smooth and one-to-one $F : \tilde{\mathcal{N}} \to \mathcal{N} \subset \mathcal{M}$ with $F$ everywhere of maximal rank and $\tilde{\mathcal{N}}$ an $n$-dimensional manifold.

**Example 1.5.** $\mathcal{M} = \mathbb{R}^3$ consider the parametrized curve $\phi(t) = (\cos(t), \sin(t), t)$ (a circular elix), $\phi$ is one-to-one and $\dot{\phi} = (-\sin(t), \cos(t), 1)$ is never 0 so the maximal rank condition is satisfied and the elix is an immersed submanifold of $\mathbb{R}^3$.

### 1.2 Vector fields

A tangent vector to a manifold $\mathcal{M}$ at a point is the tangent to a smooth curve passing through the point: given $x \in \mathcal{M}$ and $\phi(t) \in \mathcal{M}$ the curve such that $\phi(0) = x$ then

$$v_x := \left. \frac{d}{dt} \phi(t) \right|_{t=0}.$$

**Definition 1.6.** The tangent space to a $m$-dimensional manifold $\mathcal{M}$ at the point $x$ is the vector space of dimension $m$ formed by the collection of the tangent vectors at $x$ and is denoted by $T_x \mathcal{M}$.

$$T_x \mathcal{M} := \{ v = \left. \frac{d}{dt} \phi(t) \right|_{t=0} \text{, s.t. } \phi(t) \in \mathcal{M}, \forall t, \phi(0) = x \}.$$

The definition of the tangent space of a sub-manifold of $\mathbb{R}^n$ is also given in [HLW] chapter IV.5.

The tangent bundle

$$T\mathcal{M} = \bigcup_{x \in \mathcal{M}} T_x \mathcal{M}$$

is the collection of all tangent spaces, it can be given the structure of a manifold of dimension $2m$.

The tangent bundle to the circle can be identified with the cartesian product of the circle with $\mathbb{R}$, $TS^1 \simeq S^1 \times \mathbb{R}$. The tangent bundle to the sphere $TS^2$ can NOT be identified with the cartesian product of the sphere and $\mathbb{R}^2$.

A vector field on $\mathcal{M}$ is a section of the tangent bundle of $\mathcal{M}$, i.e. is a smoothly varying assignment of tangent vectors: $v : \mathcal{M} \to T\mathcal{M}$ such that
\( \mathbf{v}(x) = \mathbf{v}|_x \in T_x \mathcal{M} \). In local coordinates

\[
\mathbf{v}(x) = \sum_{i=1}^{m} \xi^i(x) \frac{\partial}{\partial x^i},
\]

\( \xi^i(x) \) are smooth functions and \( \frac{\partial}{\partial x^i} \) denote a basis of the tangent space \( T_x \mathcal{M} \).

A curve \( \phi : \mathbb{R} \to \mathcal{M} \) is an integral curve of the vector field \( \mathbf{v} \) if, when \( \phi(t) = x \), the tangent to the curve at \( t \) coincides with the vector field at \( x \), i.e. \( \dot{\phi}(t) = \mathbf{v}(x) \). This means that in local coordinates

\[
\frac{dx^i}{dt} = \xi^i(x), \quad x^i = \phi_i(t).
\]

**Example 1.7.** We consider a vector field on \( \mathbb{R}^2 \),

\[
\mathbf{v}(x, y) = y \partial_x - x \partial_y,
\]

\[
\mathbf{v}(x, y) = \begin{pmatrix} \xi^1(x, y) \\ \xi^2(x, y) \end{pmatrix} = \begin{pmatrix} y \\ -x \end{pmatrix},
\]

and to find the integral curve one has to solve

\[
\dot{x} = y, \\
\dot{y} = -x,
\]

obtaining

\[
x(t) = \cos(t)x_0 + \sin(t)y_0, \\
y(t) = -\sin(t)x_0 + \cos(t)y_0.
\]

If \( \phi(t) \) is a maximal integral curve of the vector field we denote it with

\[
\phi(t) = \exp(t\mathbf{v})x_0, \quad x_0 = \phi(0),
\]

\( \exp(t\mathbf{v})x_0 \) is the flow generated by the vector field \( \mathbf{v} \), while \( \mathbf{v} \) is called the infinitesimal generator of the flow. This notation is justified by some fundamental properties of the flow resembling known properties of the exponential mapping:

\[
\exp(t\mathbf{v}) \exp(s\mathbf{v})x_0 = \exp((t+s)\mathbf{v})x_0, \quad \exp(0\mathbf{v})x_0 = x_0,
\]

\[
\exp(t\mathbf{v})^{-1}x_0 = \exp(-t\mathbf{v})x_0, \quad \frac{d}{dt} \exp(t\mathbf{v})x_0 = \mathbf{v}|_{\exp(t\mathbf{v})x_0}
\]
and also
\[ \mathbf{v}|_{x_0} = \frac{d}{dt} \exp(t\mathbf{v})x_0 \bigg|_{t=0}, \quad \forall x_0 \in \mathcal{M}, \]
i.e. given the flow starting from \( x_0 \) we can retrieve the vector field at \( x_0 \) by differentiating the flow with respect to \( t \) and then setting \( t = 0 \). The flow of a vector field can be expanded as
\[ \exp(t\mathbf{v})x_0 = x_0 + t \mathbf{v}|_{x_0} + \mathcal{O}(t^2). \]
If \( x_0 \) is such that \( \mathbf{v}|_{x_0} = 0 \) we say that \( x_0 \) is a singularity or equilibrium point of the vector field, and this implies
\[ \exp(t\mathbf{v})x_0 = x_0, \quad \forall t. \]
Points that are not equilibrium points are called regular.

Vector fields can operate on functions as derivations.
A derivation is a linear operator defined on an algebra\(^1\) \( \mathcal{A} \), \( D: \mathcal{A} \rightarrow \mathcal{A} \)
satisfying the Leibniz rule, \( D(ab) = D(a)b + aD(b) \), for all \( a, b \in \mathcal{A} \), where \( ab \) is the product of \( a \) and \( b \) in \( \mathcal{A} \).

Given \( f: \mathcal{M} \rightarrow \mathbb{R} \) the result of applying \( \mathbf{v} \) as a derivation on \( f \) is a new function \( \mathbf{v}(f) \) such that
\[ \mathbf{v}(f(x)) = \sum_{i=1}^{m} \xi^i(x) \frac{\partial f}{\partial x^i} = \left. \frac{d}{dt} f(\exp(t\mathbf{v})x) \right|_{t=0}, \]
\( \mathbf{v}(f) \) determines the infinitesimal change of \( f \) along the flow of \( \mathbf{v} \). It is easy to verify that \( \mathbf{v} \) acts as a derivation
1. \( \mathbf{v}(\lambda f + \mu g) = \lambda \mathbf{v}(f) + \mu \mathbf{v}(g), \)
2. \( \mathbf{v}(fg) = f\mathbf{v}(g) + g\mathbf{v}(f). \)

The Lie series expansion of \( f: \mathcal{M} \rightarrow \mathbb{R} \) is an expansion of \( f \) evaluated along the flow of \( \mathbf{v} \)
\[ f(\exp(t\mathbf{v})x) = f(x) + t\mathbf{v}(f(x)) + \frac{1}{2}t^2\mathbf{v}(\mathbf{v}(f(x))) + \ldots, \]
\(^1\)An algebra is a vector space equipped with a multiplication operation, \( \cdot: \mathcal{A} \times \mathcal{A} \rightarrow \mathcal{A} \). This operation is distributive with respect to the addition of the vector space and is compatible with the product by scalars in an appropriate sense.
converging for \( t \) sufficiently near 0. This is a way of reconstructing \( f \) along the flow of \( v \) given \( v \).

The **Lie bracket** of vector fields is an operation on the set of vector fields, given two vector fields \( v \) and \( w \), \([v, w]\) is also a vector field. Such vector field is identified by the way it is acting on smooth functions, i.e. for all smooth \( f : \mathcal{M} \to \mathbb{R} \),

\[
[v, w](f) = v(w(f)) - w(v(f)).
\]

In coordinates, assuming

\[
v = \sum_{i=1}^{m} \xi^i \frac{\partial}{\partial x^i}, \quad w = \sum_{i=1}^{m} \eta^i \frac{\partial}{\partial x^i},
\]

we obtain

\[
[v, w] = \sum_{i=1}^{m} \sum_{j=1}^{m} \xi^i \frac{\partial \eta^j}{\partial x^i} \frac{\partial}{\partial x^j} - \sum_{i=1}^{m} \sum_{j=1}^{m} \eta^i \frac{\partial \xi^j}{\partial x^i} \frac{\partial}{\partial x^j}
= \sum_{j=1}^{m} \left( \sum_{i=1}^{m} \xi^i \frac{\partial \eta^j}{\partial x^i} - \sum_{i=1}^{m} \eta^i \frac{\partial \xi^j}{\partial x^i} \right) \frac{\partial}{\partial x^j}.
\]

One can verify that the following important properties hold for the Lie bracket of vector fields:

1. **bilinearity**: \( \lambda_1 v_1 + \lambda_2 v_2, w \) = \( \lambda_1 [v_1, w] + \lambda_2 [v_2, w] \).
2. **skew-symmetry**: \([v, w] = -[w, v] \).
3. **Jacobi identity**: \([v, [w, u]] + [u, [v, w]] + [w, [u, v]] = 0 \).

The **derivative map** or **differential** of a given map \( F : \mathcal{M} \to \mathcal{N} \) is a map

\[
dF : T_\mathcal{M} \to T_\mathcal{N}, \quad \text{s.t.} \quad dF|_x : T_x \mathcal{M} \to T_{F(x)} \mathcal{N}
\]

which is such that for any curve \( \phi(t) \) such that \( \phi(t)|_{t=0} = x \) and correspondingly \( F(\phi(t))|_{t=0} = F(x) \), with tangent vectors \( v_x := \frac{d}{dt} \phi(t)|_{t=0} \) and \( w_{F(x)} := \frac{d}{dt} F(\phi(t))|_{t=0} \) we have

\[
dF|_x (v_x) = w_{F(x)}.
\]

The differential is a linear map and in coordinates it is represented by the Jacobian of \( F \). Only when \( F \) is one-to-one \( dF \) maps vector fields to vector fields, [Ol].

5
Assume $F$ is one-to-one and $\mathbf{v}$ a vector field on $\mathcal{M}$ and $dF(\mathbf{v})$ a vector field on $\mathcal{N}$, one can prove that

$$F(\exp(t\mathbf{v})x) = \exp(tdF(\mathbf{v}))F(x).$$

If $\mathbf{w}$ is also a vector field on $\mathcal{M}$ than one can also prove that the Lie bracket of vector fields is invariant under $dF$, i.e.

$$dF([\mathbf{v}, \mathbf{w}]) = [dF(\mathbf{v}), dF(\mathbf{w})].$$

### 1.3 Lie groups

A Lie group is a manifold $G$ equipped with a smooth product operation "·" which gives to $G$ a group structure, e.g. there exists an identity element $e \in G$ and for any $g \in G$ its inverse $g^{-1}$ is in $G$.

Here follows a list of examples of Lie groups,

- $(\mathbb{R}, +)$
- $\text{GL}(n) = \{ A \in M^{n \times n} | \det(A) \neq 0 \}$ with the product between $n \times n$ matrices as group product,
- $\text{SL}(n) = \{ A \in M^{n \times n} | \det(A) = 1 \}$ with the product between $n \times n$ matrices as group product,
- $\text{SO}(n) = \{ A \in M^{n \times n} | \det(A) = 1, A^T A = I \}$ with the product between $n \times n$ matrices as group product,
- $\text{SP}(2r) = \{ A \in M^{2r \times 2r} | A^T J A = J \}$ with the product between $2r \times 2r$ matrices as group product,

here $M^{n \times n}$ is the set of $n \times n$ real matrices.

### 1.4 Transformation groups

A transformation group acting on a smooth manifold $\mathcal{M}$ is a Lie group $G$ and a smooth map $\Lambda : G \times \mathcal{M} \to \mathcal{M}$ such that

- $\Lambda(e, x) = x$ for all $x \in \mathcal{M}$.
- $\Lambda(g, \Lambda(h, x)) = \Lambda(g \cdot h, x)$ for all $x \in \mathcal{M}$ and $g, h \in G$.

$\Lambda$ is called a Lie group action. We say that the Lie group action is global when $\Lambda(g, x)$ is defined for all $x \in \mathcal{M}$ and $g \in G$ and local if it is defined on an open subset $\mathcal{V} \subset G \times \mathcal{M}$ such that $\{e\} \times \mathcal{M} \subset \mathcal{V}$.

Some examples:
• $\text{GL}(n, \mathbb{R})$ (or any of its subgroups) acting on $\mathbb{R}^n$ by matrix-vector multiplication.

• Any Lie group can act on itself by the group multiplication.

The set 

$$O_x = \{ m \in M \mid m = \Lambda(g, x), \ g \in G \}$$

is called orbit of the Lie group action.

**Example 1.8.** Consider the group $O(2)$ acting on $\mathbb{R}^2$ the orbits are circles around the origin of $\mathbb{R}^2$. Analogously for $O(n)$ acting on $\mathbb{R}^n$ the orbits are spheres:

$$\{ x \in \mathbb{R}^n \mid \|x\| = C \}$$

with $C$ a constant.

A lie group action is said to be transitive when there is only one orbit, 

$$O_x = M,$$

i.e.

$$\forall y \in M \exists g \in G, \ s.t. \Lambda(g, x) = y.$$ 

**Example 1.9.** The action of a Lie group $G$ on itself by left multiplication is transitive.

### 1.5 Homogeneous spaces

Given a Lie group $G$ and a subgroup $H$ we can define an equivalence relation on $G$:

$$g \sim \tilde{g} \Leftrightarrow \exists \tilde{h} \in H \text{ s.t.} \tilde{g} = g\tilde{h}.$$ 

The equivalence classes, $[g] = g \cdot H$, are called left-cosets

$$[g] = \{ gh \mid h \in H \}.$$

One can prove that if $H$ is a closed subgroup then the quotient $G/H$ (i.e. $G/\sim$) is a manifold called homogeneous space.

Recall that for $G/H$ to be a group $H$ needs to be a normal subgroup i.e. $gHg^{-1} = H$ for all $g \in G$.

In a homogeneous space the action $\Lambda : G \times G/H \rightarrow G/H$, $\Lambda(g, [\tilde{g}]) = [g\tilde{g}]$ is transitive. In fact for any $[g_1]$ and $[g_2]$ in $G/H$ it exists $g \in G$ such that $g = g_2g_1^{-1}$ and $\Lambda(g, [g_1]) = [g_2]$.

**Example 1.10.** **Relevant homogeneous spaces.** Prove that the sphere is an homogeneous space $S^2 = \text{SO}(3)/\text{SO}(2)$.
More in general $\text{SO}(n)/\text{SO}(p)$ for $p < n$ is another interesting homogeneous space called Stiefel manifold and can be identified with the set of all $n \times p$ matrices with $p$ orthonormal columns.

Analogously $\text{O}(n)/(\text{O}(p) \times \text{O}(n-p))$ is the homogeneous space also known as Grassmann manifold.

**Definition 1.11.** Given $x \in \mathcal{M}$ and $\Lambda$ a Lie group action on $\mathcal{M}$ the isotropy subgroup of $x \in \mathcal{M}$ is

$$G_x = \{ g \in G | \Lambda(g, x) = x \}.$$ 

Recall that if $H \subset G$ is a subgroup of $G$ (a Lie group) and $H$ is topologically closed then $H$ is a Lie subgroup see chapter II in [Ol].

This implies that $G_x$ is a Lie subgroup.

**Theorem 1.12.** A Lie group $G$ acts globally and transitively on $\mathcal{M}$ if and only if $\mathcal{M} \simeq G/H$ is isomorphic to the homogeneous space obtained as $G/H$ with $H = G_x$ the isotropy subgroup of any chosen $x \in \mathcal{M}$.

So any transitive Lie group action corresponds to a homogeneous space and vice-versa.

### 1.6 Lie algebra of a Lie group

A Lie algebra $\mathfrak{g}$ is a vector space with a bracket operation:

- $[\cdot, \cdot] \colon \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ is a bilinear map
- is skew-symmetric: $[u, v] = -[v, u], \forall u, v \in \mathfrak{g}$
- satisfies the Jacobi identity: $[u, [v, w]] + [w, [u, v]] + [v, [w, u]] = 0, \forall u, v, w \in \mathfrak{g}$

The Lie algebra of a given Lie group $G$ is the tangent space at the identity element $e$, $\mathfrak{g} := T_e G$.

This means that $\mathfrak{g}$ is a linear vector space, for matrix Lie groups the Lie algebra is typically a linear vector subspace of $M^{n \times n}$.

One can verify that

- the Lie algebra of $(\mathbb{R}, +)$ is $\mathbb{R}$
- the Lie algebra of $\text{GL}(n)$ is $\mathfrak{gl}(n) = M^{n \times n}$,
- the Lie algebra of $\text{SL}(n)$ is $\mathfrak{sl}(n) = \{ A \in M^{n \times n} | \text{trace}(A) = 0 \}$,
• the Lie algebra of \( \text{SO}(n) \) is \( \mathfrak{so}(n) = \{ A \in M^{n \times n} | A^T + A = 0 \} \),

• the Lie algebra of \( \text{Sp}(2r) \) is \( \mathfrak{sp}(2r) = \{ A \in M^{2r \times 2r} | AJ + JA^T = 0 \} \),

Proof. Consider \( A(t) \in \text{Sp}(2r) \) \( A(0) = I \), we get \( V = \frac{d}{dt} A(t) \bigg|_{t=0} \in \mathfrak{sp}(2r) \). We differentiate \( A(t)JA(t)^T = J \) and obtain

\[ \dot{A}JA^T + AJ\dot{A}^T = 0. \]

Setting \( t = 0 \) we obtain

\[ VJ + JV^T = 0. \]

We now consider an alternative definition of the Lie algebra of a Lie group, we will see how this is equivalent to the definition given above, and will naturally obtain the definition of the bracket operation on \( \mathfrak{g} \).

Consider the left multiplication of the Lie group \( L_g : G \to G, L_g(\tilde{g}) = g\tilde{g} \), consider also the derivative mapping or differential of \( L_g \), \( dL_g : T_xG \to T_{gx}G \).

Definition 1.13. A vector field on \( G, v \), is left invariant if

\[ dL_g(v) = v. \]

Since for a left invariant vector field \( v(e) = A \) implies \( dL_g(v(e)) = v(g) \), once we know how the vector field looks like at the identity, \( e \), we know how it looks like everywhere. For this reason we can identify the tangent space at the identity of the Lie group \( G \), i.e. the Lie algebra \( \mathfrak{g} \), with the set of left invariant vector fields. An analogous definition and identification can be given for right invariant vector fields.

The Lie algebra can be also used to describe the tangent space to \( G \) at any point. Here is the case of the orthogonal group.

Example 1.14. Consider \( \gamma(t) \in O(n) \). \( \gamma(t)^T\gamma(t) = I \) assume \( \gamma(0) = Q \) and \( \dot{\gamma}(0) = W \). By differentiating with respect to \( t \) and setting \( t = 0 \) we obtain

\[ W^TQ + Q^TW = 0. \]

Set \( A := Q^TW \) and substitute \( W = QA \) in the previous equation, obtaining \( A^T + A = 0 \). So we obtain a characterization of the tangent space of \( O(n) \) at \( Q \) by means of \( \mathfrak{so}(n) \):

\[ T_Q O(n) = \{ W = QA | A \in \mathfrak{so}(n) \}. \]

Analogous results can be obtained for other matrix Lie groups.
Recall from chapter I in [Ol]: given a mapping \( F : M \to N \) the derivative mapping \( dF \) does not map vector fields to vector fields unless \( F \) is one-to-one. Assume \( F \) is one-to-one and \( \mathbf{v} \) a vector field on \( M \) and \( dF(\mathbf{v}) \) a vector field on \( N \), one can prove that
\[
F(\exp(t\mathbf{v})x) = \exp(tdF(\mathbf{v}))F(x). \tag{2}
\]
If \( \mathbf{w} \) is also a vector field on \( M \) than one can also prove that the Lie bracket of vector fields is invariant under \( dF \), i.e.
\[
dF([\mathbf{v}, \mathbf{w}]) = [dF(\mathbf{v}), dF(\mathbf{w})].
\]
If \( M = G \) (a Lie group) the right multiplication \( R_g : G \to G \) is one-to-one. Assume \( \mathbf{v} \) and \( \mathbf{w} \) are right invariant vector fields: \( dR_g(\mathbf{v}) = \mathbf{v} \) and \( dR_g(\mathbf{w}) = \mathbf{w} \), then
\[
dR_g([\mathbf{v}, \mathbf{w}]) = [dR_g(\mathbf{v}), dR_g(\mathbf{w})] = [\mathbf{v}, \mathbf{w}],
\]
so that also the Lie bracket \([\mathbf{v}, \mathbf{w}]\) of the two right invariant vector fields is right invariant. The Lie algebra is closed under the Lie bracket of vector fields. So the Lie bracket of vector fields is the bracket operation of the Lie algebra \( g \) of a Lie group \( G \). Recall that the Lie bracket of vector fields is bilinear skew-symmetric and satisfies the Jacobi identity, see [Ol] chapter I.

**Example 1.15.** Consider \( \mathfrak{gl}(n) \) and the bracket of vector fields by a similar argument to that used in example 1.14 we obtain that a left invariant vector field \( \mathbf{v}_A \) on \( \text{GL}(n) \) has coordinates given by the matrix product \( XA \) with \( A \in \mathfrak{gl}(n) \) and \( X \in \text{GL}(n) \). Consider two left invariant vector fields \( \mathbf{v}_A \) and \( \mathbf{v}_B \) and write them as derivation operators:
\[
\mathbf{v}_A(X) = \sum_{i,j,k=1}^{n} x_{i,k} a_{k,j} \frac{\partial}{\partial x_{i,j}}, \quad \mathbf{v}_B(X) = \sum_{i,j,k=1}^{n} x_{i,k} b_{k,j} \frac{\partial}{\partial x_{i,j}},
\]
here we denote with \( a_{i,j} = (A)_{i,j} \) the \((i,j)\)-element of the matrix \( A \). Computing the Lie bracket of the two vector fields we obtain
\[
[\mathbf{v}_A, \mathbf{v}_B](X) = \sum_{i,j,k=1}^{n} x_{i,k} a_{k,j} \sum_{s=1}^{n} b_{j,s} \frac{\partial}{\partial x_{i,s}} - \sum_{l,r,s=1}^{n} x_{l,r} b_{r,s} \sum_{j=1}^{n} a_{s,j} \frac{\partial}{\partial x_{l,j}}
= \sum_{i,k,s=1}^{n} x_{i,k} (AB - BA)_{k,s} \frac{\partial}{\partial x_{i,s}}.
\]
The Lie bracket of the two vector fields is a vector field with coordinates the matrix commutator of \( A \) and \( B \): \([A, B] = AB - BA \). The situation is analogous for the case of right invariant vector fields a part from a change of sign.
1.7 The exponential map

Consider \( v \) a right invariant vector field on \( G \) and the right multiplication of the Lie group \( R_g \). Using (2) we obtain

\[
R_g(\exp(tv)e) = \exp(tdR_g(v))g
\]

and further using the right invariance of \( v \) on the right hand side we get

\[
(\exp(tv)e)g = \exp(tv)g,
\]

so that the left multiplication of \( g \) by the flow through \( e \) of \( v \) is equal to the flow through \( g \) of \( v \). We therefore can identify the flow of the right invariant vector field to be the corresponding one parameter subgroup\(^2\) of \( G \):

\[
\exp(tv) := \exp(tv)e.
\]

We can also define the exponential map \( \exp : g \to G \) as

\[
v \in g \mapsto \exp(tv)e|_{t=1} \in G.
\]

**Example 1.16.** The flow of a right invariant vector field

\[
v_A = \sum_{i,j,k} a_{i,j} x_{k,j} \frac{\partial}{\partial x_{i,j}},
\]

is \( \gamma(t) \) such that \( \dot{\gamma} = v_A(\gamma(t)) \), in coordinates

\[
\dot{\gamma}_{ij} = \sum_{i,j=1}^{n} \left( \sum_{k=1}^{n} a_{i,k} \gamma_{kj}(t) \right), \quad \text{i.e.} \quad \dot{\gamma} = A \gamma,
\]

and \( \gamma(t) = \exp(tA)\gamma(0) \). For the left invariant vector fields the flow is instead of the type \( \eta(t) = \eta(0) \exp(tA) \).

**Exercise 1.17.** Show that \( \exp(O) = e \), where \( O \) is the zero element in \( g \) and \( e \) is the identity element of \( G \). Show that the derivative mapping of \( \exp \) at \( O \) is the identity mapping in \( g \).

The results of the previous exercise guarantee that \( \exp \) is a local diffeomorphism from a neighborhood of \( O \in g \) to a neighborhood of \( e \in G \). This follows from the inverse function theorem. (See also [HLW] chapter IV.6 on this topic.)

The exponential mapping can be used to put local coordinates on the Lie group by means of the Lie algebra.

\(^2\)One parameter subgroup: a subgroup depending on one parameter, in this case \( t \).
Theorem 1.18. Let $G$ be a connected Lie group with Lie algebra $\mathfrak{g}$. Every group element can be written as a product of exponentials:

$$g = \exp(V_1) \exp(V_2) \cdots \exp(V_k),$$

for $V_1, \ldots, V_k \in \mathfrak{g}$.

1.8 Some properties of the exponential in matrix Lie groups

We want to consider

$$\frac{d}{dt} \exp(\sigma(t)) e \bigg|_{t=0},$$

where $\sigma(t)$ is a curve in $\mathfrak{gl}(n)$. We proceed giving two Lemmas which are used for this aim.

Lemma 1.19. Variation of constants formula.

The solution of the differential equation

$$\dot{u} = Au + w, \quad u(0) = u_0,$$

where $A$ is a $m \times m$ constant matrix and $u_0, w \in \mathbb{R}^m$ are fixed, is

$$u(t) = e^{tA}u_0 + \int_0^t e^{(t-x)A}wdx.$$

Proof. To find the solution of the considered differential equation we use the integrating factor $e^{-xA}$, we obtain

$$e^{-xA}\dot{u}(x) - Ae^{-xA}u(x) = e^{-xA}w.$$

We now integrate between 0 and $t$ and obtain

$$e^{-tA}u(t) - u(0) = \int_0^t e^{-xA}wdx,$$

and multiplying on both sides with $e^{tA}$ we obtain the result.

Corollary 1.20. If $w \in \mathbb{R}^m$ and $A$ is a $m \times m$ matrix we have that

$$\int_0^t e^{(t-x)A}wdx = \frac{e^{tz} - 1}{z} \bigg|_{z=A} w.$$
Proof. We expand the integral at the left hand side of the equality by using the Taylor series of the exponential mapping and we obtain
\[
\int_0^t e^{(t-x)A}wdx = \sum_{i=0}^{\infty} \int_0^t \frac{(t-x)^i}{i!} A^iwdx,
\]
and since
\[
\int_0^t \frac{(t-x)^i}{i!}wdx = \frac{t^{i+1}}{(i+1)!}w,
\]
we obtain
\[
\int_0^t e^{(t-x)A}wdx = \sum_{i=0}^{\infty} A^i \frac{t^{i+1}}{(i+1)!}w = \sum_{k=1}^{\infty} A^{k-1} \frac{t^k}{k!}w.
\]
Now one can verify that
\[
\frac{e^{tz} - 1}{z} = \sum_{k=1}^{\infty} \frac{z^{k-1} t^k}{k!},
\]
(use the expansion for \(e^{tz}\)), which implies that
\[
\int_0^t e^{(t-x)A}wdx = \frac{e^{tz} - 1}{z} \bigg|_{z=A} w.
\]

This Lemma is used in the proof of the next Lemma.

Lemma 1.21. Assume \(\sigma(t)\) is a \(n\times n\) matrix for each \(t\) then we have that
\[
\left( \frac{d}{dt} e^{\sigma(t)} \right) e^{-\sigma(t)} = \frac{e^z - 1}{z} \bigg|_{z=ad_{\sigma}} (\dot{\sigma}),
\]
where for two \(n\times n\) matrices \(B\) and \(C\) we have \(ad_B(C) = [B, C] = BC - CB\), where \([\cdot, \cdot]\) is the matrix commutator.

Proof. Consider \(B(s, t) = \left( \frac{d}{dt} e^{\sigma(t)} \right) e^{-\sigma(t)}\). By differentiating with respect to \(s\) we obtain
\[
\frac{\partial}{\partial s} B(s, t) = \left( \frac{d}{dt} (\sigma(t)e^{\sigma(t)}) \right) e^{-\sigma(t)} - \left( \frac{d}{dt} e^{\sigma(t)} \right) e^{-\sigma(t)} \sigma(t)
= \dot{\sigma}(t)e^{\sigma(t)}e^{-\sigma(t)} + \sigma(t) \left( \frac{d}{dt} e^{\sigma(t)} \right) e^{-\sigma(t)} - B(s, t)\sigma(t)
= \dot{\sigma}(t) + [\sigma(t), B(s, t)].
\]

13
This means that
\[ \frac{\partial}{\partial s} B(s, t) = \text{ad}_\sigma(B) + \dot{\sigma}, \]
and we have \( B(0, t) = O \). Note that \( \text{ad}_\sigma \) is a linear operator acting on \( n \times n \) matrices, and can be represented as a \( n^2 \times n^2 \) matrix. Then taking \( A = \text{ad}_{\sigma(t)} \), in Lemma 1.19 and Corollary 1.20 we have
\[ B(s, t) = e^{sz} \left. \frac{1}{z} \right|_{z = \text{ad}_{\sigma(t)}} (\dot{\sigma}(t)). \]

From Lemma 1.21 we have that
\[ \frac{d}{dt} e^{\sigma(t)} = e^{\sigma - 1} \left. \frac{\dot{\sigma}}{z} \right|_{z = \text{ad}_\sigma} (\dot{\sigma}) \cdot e^{\sigma(t)}, \]
and for ease of notation we define
\[
\text{dexp}_{\sigma(t)} (\dot{\sigma}(t)) := \left. \frac{e^{\sigma - 1}}{z} \right|_{z = \text{ad}_\sigma} (\dot{\sigma}) = \sum_{k=1}^{\infty} \frac{1}{k!} \text{ad}^{k-1}_\sigma (\dot{\sigma}(t)) = \dot{\sigma}(t) + \frac{1}{2} [\sigma(t), \dot{\sigma}(t)] + \frac{1}{3!} [\sigma(t), [\sigma(t), \dot{\sigma}(t)]] + \ldots
\]

2 Integration methods on manifolds

2.1 Introduction and motivation

We are interested in deriving intrinsic numerical integration methods for the problem
\[ \dot{y} = F(y) \quad (4) \]
\[ y(t_0) = y_0 \quad (5) \]
with \( y_0 \in \mathcal{M} \), \( \mathcal{M} \) a smooth manifold, and \( F \) a vector field on \( \mathcal{M} \), i.e. \( F(y(t)) \in T_{y(t)}\mathcal{M} \) for all \( t \). Using a classical Runge-Kutta or multi-step method to approximate this problem does not automatically produce approximations on \( \mathcal{M} \). Our aim is to design numerical methods which by construction produce approximations on the manifold. We call them intrinsic because we use only operations which make sense in the manifold setting.
As an example consider the following differential equation on the orthogonal group
\[ \dot{Y} = A(Y) \cdot Y, \quad Y(0) = Y_0, \quad (6) \]
where \( Y \) and \( A(Y) \) are \( n \times n \) matrices, \( A(Y) \) is skew-symmetric for all \( Y \) and \( Y_0 \) is an orthogonal matrix. The solution of (6) is an orthogonal matrix in fact if we take the derivative w.r.t. time of \( Y(t)^T Y(t) \) we obtain
\[
\frac{d}{dt} Y(t)^T Y(t) = \dot{Y}^T Y + Y^T \dot{Y} = Y^T A(Y)^T Y + Y^T A(Y) Y
\]
\[
= -Y^T A(Y) Y + Y^T A(Y) Y = 0,
\]
which means that \( Y(t)^T Y(t) \) is constant and therefore
\[ Y(t)^T Y(t) = Y_0^T Y_0 = I, \quad \forall t, \]
i.e. \( Y(t) \) is an orthogonal matrix for all \( t \). The format (6) is a consequence of the characterization of the tangent space discussed in example 1.14 and is valid in general for \( A(Y) \) belonging to the Lie algebra \( g \) of a Lie group \( G \) and \( Y_0 \in G \).

This type of differential equations arise in rigid body dynamics, structural mechanics, and many other fields of science and engineering. Often it is important to compute numerical approximations of the solution which belong to the Lie group.

**Example 2.1. The rigid body equations** Euler’s theorem states that the general displacement of a rigid body (or coordinate frame) with one point fixed is a rotation about some axis. This leads to Euler’s equations for the free rigid body,
\[ \dot{\pi} = \text{skew}(T^{-1} \pi), \quad (7) \]
where
\[
\text{skew}(v) = \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix},
\]
and
\[
T = \begin{pmatrix} I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_3 \end{pmatrix}.
\]

Here \( \pi \) is the angular momentum, \( I_1, I_2, I_3 \) are the principal moments of inertia of the body. According to Euler’s theorem there exists a \( 3 \times 3 \) rotation
matrix $Q(t)$ such that $\pi = Q\pi_0$. Such a matrix must then satisfy the following differential equation

$$\dot{Q} = \text{skew}(T^{-1}\pi) Q. \quad (8)$$

Correct and efficient numerical integration of equations (7) and (8) is of interest for applications in molecular dynamics and celestial mechanics, for example.

Suppose we apply the Forward Euler’s method to (6), i.e.

$$Y_{n+1} = Y_n + hA(Y_n)Y_n, \quad n = 0, 1, 2\ldots$$

and assume $Y_0^TY_0 = I$. We want to know if this implies that also $Y_1^TY_1 = I$. By direct calculation we obtain

$$Y_1^TY_1 = (Y_0^T + hY_0^TA(Y_0)^T)(Y_0 + hA(Y_0)Y_0)$$

$$= I + hY_0^TA(Y_0)Y_0 + hY_0^TA(Y_0)^TY_0 + h^2Y_0^TA(Y_0)^TAY_0Y_0$$

$$= I - h^2Y_0^TA(Y_0)^2Y_0,$$

in general we can not expect the term $Y_0^TA(Y_0)^2Y_0$ to vanish and therefore we can not expect $Y_1$ to be orthogonal.

We want to find strategies alternative to the forward Euler’s method which maintain the orthogonality under time discretization. Given a $n \times n$ matrix $B$ with constant entries we know that

$$\exp(tB) = e^{tB} = \sum_{k=0}^{\infty} \frac{t^k}{k!} B^k$$

is the solution of the matricial differential equation

$$\dot{Y} = BY, \quad Y(0) = I.$$

Observe that if $B$ is a skew-symmetric matrix the above equation is a special case of equation (6) and $\exp(tB)$ is an orthogonal matrix. We consider the following alternative to the forward Euler method for the numerical integration of equation (6),

$$Y_{n+1} = \exp(hA(Y_n))Y_n. \quad (9)$$

Provided $Y_0^TY_0 = I$ and exploiting the orthogonality of $\exp(hA(Y_0))$ we can verify that $Y_1$ is an orthogonal matrix, in fact

$$Y_1^TY_1 = Y_0^T \exp(hA(Y_0))^T \exp(hA(Y_0))Y_0 = I,$$
in other words the proposed method produces numerical approximations to
the solution of (6) which belong to the set of orthogonal matrices.

The method (9) is known as Lie-Euler method and has order 1. In fact
if we consider the Taylor expansion of $Y(h)$ and $Y_1(h)$ around zero, we
can easily verify that they coincide up to second order in $h$. We have

\[ Y(h) = Y_0 + h A(Y_0) Y_0 + \frac{h^2}{2!} \dot{Y}(0) + \ldots \]

and

\[ Y_1(h) = Y_1(0) + h \dot{Y}_1(0) + \frac{h^2}{2!} \ddot{Y}_1(0) + \ldots \]

Now $Y_1(0) = Y_0$ and since

\[ \left. \frac{d}{dh} Y_1(h) \right|_{h=0} = \left. \frac{d}{dh} \exp(h A(Y_0)) Y_0 \right|_{h=0} = A(Y_0) \exp(h A(Y_0)) Y_0 \big|_{h=0} = A(Y_0) Y_0, \]

we easily see that the two Taylor expansions coincide up to terms of order
at least 2 in $h$.

In the next sections we will generalize this method following two different
strategies which lead to two different classes of methods.

### 2.2 Methods based on frame vector fields

**Definition 2.2.** A set of vector fields $\{E_1, \ldots, E_d\}$ on the manifold $M$ of
dimension $m \leq d$ is a set of frame vector fields if

\[ T_x M = \text{span}\{E_1|_x, \ldots, E_d|_x\}, \quad \forall x \in M. \]

Given any vector field $F$ on $M$ we have

\[ F(y) = \sum_{i=1}^{d} f_i(y) E_i(y). \]

**Definition 2.3.** We denote with $F_p$ the vector field

\[ F_p(x) = \sum_{i=1}^{d} f_i(p) E_i(x) \]

we say that $F_p$ is the vector field $F$ frozen at the point $p$.

Given at $M$ is a manifold with a set of frame vector fields we can define
intrinsic Runge-Kutta like methods as follows:

**Commutator-free method**
for \( r = 1 : s \) do 
\[ Y_r = \exp(\sum_{k=1}^{s} \alpha_{r,j}^k F_k) \cdots \exp(\sum_{k=1}^{s} \alpha_{r,1}^k F_k)(p) \]
\[ F_r = h F_{Y_r} = h \sum_{i=1}^{d} f_i(Y_r) \mathcal{E}_i \]
end
\[ y_1 = \exp(\sum_{k=1}^{s} \beta_{j}^k F_k) \cdots \exp(\sum_{k=1}^{s} \beta_{1}^k F_k)p \]

Here \( n \) counts the number of time steps and \( h \) is the step-size of integration. The integrator has \( s \) stages and parameters \( \alpha_{r,j}^k, \beta_{j}^k \). Each new stage value is obtained as a composition of \( J \) exponentials of linear combinations of vector fields frozen at the previously computed stage values.

In the following tableaus we report the coefficients of a method of order 3 and a method of order 4. The method of order 3 requires the computation of one exponential of each internal stage value and the composition of two exponentials for updating the solution. In the order 4 method the first three stage values require one exponential each, while the fourth stage and the solution update require two exponentials.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/2</td>
<td>1/2</td>
<td>1/2</td>
<td>0</td>
</tr>
<tr>
<td>2/3</td>
<td>0</td>
<td>2/3</td>
<td>0</td>
</tr>
<tr>
<td>1/2</td>
<td>0</td>
<td>3/4</td>
<td>0</td>
</tr>
<tr>
<td>-1/2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>-1/2</td>
<td>1/4</td>
<td>1/6</td>
<td>1/12</td>
</tr>
</tbody>
</table>

Example 2.4. Let \( \mathcal{M} \) be a manifold acted upon transitively by a Lie group \( G \). Denote with \( \Lambda : G \times \mathcal{M} \to \mathcal{M} \) the Lie group action. Suppose \( E_1, \ldots, E_d \) a basis of the Lie algebra then \( F_{E_1}, \ldots, F_{E_d} \) obtained by
\[ F_{E_i}(x) = \frac{d}{dt} \Lambda(\exp(tE_i), x) \bigg|_{t=0}, \]
are a set of frame vector fields.

In particular for matrix Lie groups consider the vector field \( A(Y)Y \) of equation (6). Here \( A(y) \in \mathfrak{g} \) and \( A(Y) = \sum_{i=1}^{d} a_i(y)E_i \) with \( E_1, \ldots, E_d \) a basis of the Lie algebra\(^3\). The vector field frozen at a point \( P \in G \) is simply \( A(P)Y \).

\(^3\)Say for \( \mathfrak{so}(n) \) a basis is given by the matrices of rank 2 of the type \( e_i e_j^T - e_j e_i^T \) with \( e_i, e_j \in \mathbb{R}^n \) canonical vectors, and \( i = 1, \ldots, n, j \leq i - 1. \)
Note that at each stage in the commutator-free methods we allow for the composition of at most $J$ exponentials. As, in general, the computation of matrix exponentials is a computationally demanding task, it is of interest to find methods in this class which require a minimal number of exponentials, see [CMO] for details.

Given the above format for the methods the challenge is to find parameters

$$\alpha_{J,i}^k, \alpha_{1,i}^k, \ i, k = 1, \ldots, s, \ \beta_{J}^k, \ldots, \beta_{1}^k, \ k = 1, \ldots, s$$

such that the formulae above produce a method of a desired order. The order theory for these methods can be developed as usual by requiring that

$$\left. \frac{d^r}{dh^r} Y_1 \right|_{h=0} = \left. \frac{d^r}{dh^r} Y(h) \right|_{h=0}, \ r = 1, \ldots, p.$$ 

The number of order conditions for order $p$ is higher than for classical Runge-Kutta methods. A complete treatment of this subject can be found in [O].

If we assume $J = i - 1$ and allow for just one of the $\alpha$ values to be different from zero in each exponential, we obtain the Crouch and Grossman methods, [CG].

**Crouch and Grossman**

for $i = 1 : s$ do

$$Y_i = \exp(a_{i,i-1}F_{i-1}) \cdots \exp(a_{i,1}F_1)Y_n$$

$$F_r = hF_{Y_r} = h \sum_{i=1}^d f_i(Y_r) \xi_i$$

end

$$Y_{n+1} = \exp(b_sF_s) \cdots \exp(b_1F_1)Y_n.$$ 

Crouch and Grossman pioneered the field of integration methods on manifolds (and Lie Group methods) in the nineties. These methods are defined by a tableau similar to the classical Runge-Kutta Butcher tableau, but also in this case the methods require extra order conditions compared to classical Runge-Kutta methods. An example of a method of order 3 is the following:

\[
\begin{array}{c|cccc}
0 & & & & \\
-\frac{1}{24} & & & & \\
\frac{17}{24} & -\frac{1}{24} & & & \\
\frac{161}{24} & \frac{16}{24} & -6 & & \\
& & 1 & -\frac{2}{3} & \frac{2}{3} \\
\end{array}
\]
In the case of matrix Lie groups, equation (6), using the results of example 2.4 and applying this method we obtain:

\[
\begin{align*}
Y_1 &= Y_n \\
Y_2 &= \exp(-h/24A(Y_1))Y_n \\
Y_3 &= \exp(-6hA(Y_2)) \exp(161/24hA(Y_1))Y_n \\
Y_{n+1} &= \exp(2/3hA(Y_3)) \exp(-2/3hA(Y_2)) \exp(hA(Y_1))Y_n,
\end{align*}
\]

this method requires 6 exponentials. If we use the commutator-free method of the same order we obtain:

\[
\begin{align*}
Y_1 &= Y_n \\
Y_2 &= \exp(h/3A(Y_1))Y_1 \\
Y_3 &= \exp(2/3hA(Y_2))Y_1 \\
Y_{n+1} &= \exp(-1/2hA(Y_1) + 3/4hA(Y_3))Y_2,
\end{align*}
\]

requiring three exponentials.

### 2.3 RK-MK methods

Assume the manifold \(\mathcal{M}\) is acted upon transitively by a Lie group \(G\). We use the action and perform the following change of variables for (4):

\[
y(t) = \Lambda(\exp(\sigma(t)), y_0), \quad \sigma(t) \in g,
\]

valid in a neighborhood of \(y_0 \in \mathcal{M}\). The idea is to derive an equation for \(\sigma\) in the Lie algebra and then to integrate this equation with a classical Runge-Kutta method and obtain an approximation \(\hat{\sigma} \approx \sigma(h)\) which still belongs to the same Lie algebra and which, via exponentiation, and the Lie group action, generates

\[
y(h) \approx y_1 = \Lambda(\exp(\hat{\sigma}), y_0).
\]

Consider the mapping \(\Lambda_{y_0} : G \to \mathcal{M}\), defined by \(\Lambda_{y_0}(g) = \Lambda(g, y_0)\). By differentiation we obtain

\[
F(y(t)) = \frac{d}{dt} y(t) = \frac{d}{dt} \Lambda(\exp(\sigma(t)), y_0) = \frac{d}{dt} \Lambda_{y_0}(\exp(\sigma(t)))
\]

and further, from the definition of differential (derivative mapping), [Ol] chapter I, we obtain

\[
F(y(t)) = d\Lambda_{y_0}(\frac{d}{dt} \exp(\sigma(t))).
\]
By assuming now without substantial loss of generality that we work with matrix Lie groups, from Lemma 1.9 we obtain

\[ F(y(t)) = d\Lambda_{y_0}(dR_{\exp(\sigma)}d\exp(\dot{\sigma})) = (d\Lambda_{y_0} \circ dR_{\exp(\sigma)})(d\exp(\dot{\sigma})), \quad (10) \]

where \( R_g \) is the right multiplication by \( g \) in the Lie group.

Using the same set of frame vector fields defined in example 2.4 in the previous section, \( F_{E_i}(x) = d\Lambda_x(E_i) \), and the linearity of the map \( d\Lambda_x \)

\[ F(x) = \sum_{i=1}^{d} f_i(x)F_{E_i}(x) = d\Lambda_x(\sum_{i=1}^{d} f_i(x)E_i), \]

here we can define \( f(x) := \sum_{i=1}^{d} f_i(x)E_i \in \mathfrak{g} \). Setting \( x = \Lambda_{y_0}(\exp(\sigma(t))) \) we obtain the following expression for the left hand side of (10)

\[ F(y(t)) = F(\Lambda_{y_0}(\exp(\sigma(t)))) = d\Lambda_{\Lambda_{y_0}(\exp(\sigma))}(f(\Lambda_{y_0}(\exp(\sigma)))). \]

Now we have that

\[ d\Lambda_{y_0} \circ dR_{\exp(\sigma)} = d\Lambda_{y_0} \circ R_{\exp(\sigma)} = d\Lambda_{\Lambda_{y_0}(\exp(\sigma))}, \]

see [Ol] chapter I. By substituting at the right hand side of (10) we obtain

\[ d\Lambda_{\Lambda_{y_0}(\exp(\sigma))}(\sum_{i=1}^{d} f_i(y(t))E_i) = d\Lambda_{\Lambda_{y_0}(\exp(\sigma))}(d\exp(\dot{\sigma})), \]

which is fulfilled if

\[ d\exp(\dot{\sigma}) = f(\Lambda_{y_0}(\exp(\sigma))), \]

and which in turn gives the differential equation for \( \sigma \) in the Lie algebra \( \mathfrak{g} \):

\[ \dot{\sigma} = d\exp_{\sigma}^{-1}(f(\Lambda_{y_0}(\exp(\sigma))). \quad (11) \]

Here \( d\exp_{\sigma} \) is an invertible map provided \( \|\sigma\| < \pi \), see [HLW] (Lemma 4.2 chap. III, 4.1 ) for details, and the inverse is given by

\[ d\exp_{\sigma}^{-1}(u) := \left. \frac{z}{e^z - 1} \right|_{z = \text{ad}_{\sigma}^{-1}} (u) = \sum_{k=0}^{\infty} \frac{B_k}{k!} \text{ad}_{\sigma}^k (u). \quad (12) \]

Here the coefficients \( B_k \) are called Bernoulli numbers, the first four of them are

\[ B_0 = 1, \quad B_1 = -\frac{1}{2}, \quad B_2 = \frac{1}{6}, \quad B_3 = 0. \]
We solve numerically with a Runge-Kutta method equation (11) in place of applying the same method directly to equation (4). The approximation $\tilde{\sigma} \approx \sigma(h)$ is then used to construct $y_1 \approx y(h)$ via exponentiation and the Lie group action, i.e.

$$y_1 = \Lambda(\exp(\tilde{\sigma}), y_0).$$

This procedure gives the Runge-Kutta Munthe-Kaas (RKMK) methods, and has been originally presented in [MK], in an equivalent but different way.

Assuming $\sigma(h) = \tilde{\sigma} + O(h^{p+1})$, (which means that the Runge-Kutta method considered is of order $p$) the local error $\|y(h) - y_1\| = \|\Lambda(e^{\tilde{\sigma}}, y_0) - \Lambda(e^\sigma, y_0)\|$ is also $O(h^{p+1})$. This can be shown using the Taylor expansion of the exponential and of the action. In other words if we apply a RK method of order $p$ to (11) we obtain an approximation of the same order for (4).

In the practical implementation of this strategy, as $\sigma(h) = O(h)$, (remember that $\sigma(0) = 0$) the infinite series defining $\text{dexp}^{-1}_\sigma$, (12), can be truncated to the right order of consistency in $h$, including just a small number of terms. We here report the algorithm for the RK-MK methods applied to (4).

**RKMK**

for $i = 1 : s$ do

$$\sigma_i = h \sum_{j=1}^{i-1} a_{i,j} \text{dexp}\_\sigma_j \left( f(\Lambda_{Y_n}(\exp(\sigma_j))) \right)$$

end

$$\tilde{\sigma} = h \sum_{i=1}^s b_i \text{dexp}\_\sigma_i \left( f(\Lambda_{Y_n}(\exp(\sigma_i))) \right)$$

$$Y_1 = \Lambda_{Y_n}(\exp(\tilde{\sigma})).$$

where we have denoted with $\text{dexp}\_\sigma^{-1}$ the truncation (to the correct consistency order) of the expansion (12). Here $a_{i,j}$ and $b_i$ $i = 1, \ldots, s$ $j = 1, \ldots, s$ are the parameters of a classical Runge-Kutta method.

Alternatively the algorithm can be written in the following format.

**RKMK**

for $i = 1 : s$ do

$$\sigma_i = h \sum_{j=1}^{i-1} a_{i,j} K_j$$

$$K_i = f(\Lambda_{Y_n}(\exp(\sigma_i)))$$

$$\tilde{K}_i = \text{dexp}\_\sigma^{-1}(K_i)$$

end

$$\tilde{\sigma} = h \sum_{i=1}^s b_i \tilde{K}_i$$

$$Y_1 = \Lambda_{Y_n}(\exp(\tilde{\sigma})).$$
We stress once more that for the RKMK methods the Runge-Kutta parameters coincide with the parameters of the classical Runge-Kutta methods and no extra order conditions are produced in this case. We can for example consider the Butcher tableau for the Heuns method of order 3, see [HNW] p. 135. This method applied to the ordinary differential equation 
\[ \dot{y} = g(y), \quad y(0) = y_0 \]
takes the following form:

\[
\begin{array}{c|cccc}
0 & & & & \\
\frac{1}{3} & 1 & & & \\
\frac{2}{3} & 2 & & & \\
\frac{1}{3} & 0 & & & \\
\end{array}
\]

\[
\begin{align*}
Y_1 &= y_n \\
Y_2 &= y_n + \frac{h}{3}g(Y_1) \\
Y_3 &= y_n + \frac{2h}{3}g(Y_2) \\
y_{n+1} &= y_n + \frac{h}{4}(g(Y_1) + 3g(Y_3)).
\end{align*}
\]

We use now this method on a matrix Lie group equation in the RK-MK fashion. The Lie group action is the action of \( G \) on itself by matrix-matrix multiplication, and the exponential is the matrix exponential. We obtain

\[
\begin{align*}
\sigma_1 &= 0 \\
K_1 &= A(Y_n) \\
\tilde{K}_1 &= \text{dexp}_O^{-1}(K_1) = K_1 \\
\sigma_2 &= \frac{h}{3}K_1 \\
K_2 &= A(\exp(\sigma_2)Y_n) \\
\tilde{K}_2 &= K_2 - \frac{1}{2}[\sigma_2, K_2] + \frac{1}{12}[\sigma_2, [\sigma_2, K_2]] \\
\sigma_3 &= \frac{2}{3}hK_2 \\
K_3 &= A(\exp(\sigma_3)Y_n) \\
\tilde{K}_3 &= K_3 - \frac{1}{2}[\sigma_3, K_3] + \frac{1}{12}[\sigma_3, [\sigma_3, K_3]] \\
\tilde{\sigma} &= \frac{h}{4}\tilde{K}_1 + \frac{3}{4}h\tilde{K}_3 \\
Y_{n+1} &= \exp(\tilde{\sigma})Y_n.
\end{align*}
\]

The described method requires the computation of 4 matrix commutators and 3 matrix exponentials per time step. It is possible to reduce the number of commutators to 1 for methods of order 3 by using techniques described in [MKO].

### 2.4 Magnus methods for linear systems of ODEs

A complete treatment of Lie group methods includes the methods based on Magnus expansion. We refer to [HLW] p.121-123, for this topic.
2.5 Further implementation details for the implementation of Lie group methods

We have seen that the Commutator-free and Crouch and Grossmann methods and RK-MK methods, applied to differential equations on manifolds acted upon by Lie groups, produce numerical approximations which belong to the manifold.

We leave as an exercise to the reader to show with similar arguments that the methods based on Magnus series are also producing approximations of linear differential equations on the Lie group $G$ which belong to $G$.

For techniques for the approximation of the matrix exponential in a Lie algebraic setting see [CI1], [CI2], [IZ].

In Lie group methods we exploit the crucial property that the exponential mapping is a local diffeomorphism from a neighborhood of $O \in \mathfrak{g}$ to put local coordinates on the Lie group $G$. This can be done also in other ways. In the case of quadratic Lie groups, i.e.

$$G := \{ y \in GL(n) | yPy^T = P \}, \quad \mathfrak{g} := \{ x \in \mathfrak{gl}(n) | xP + Px^T = O \},$$

where $P$ is a fixed $n \times n$ invertible matrix (like for example $SO(n): P = I$, $SP(2r): P = J$) it is possible to use alternatively the Cayley transformation as a coordinate map:

$$\text{cay}(A) = (I - \frac{1}{2}A)^{-1}(I + \frac{1}{2}A),$$

see for exemple [LS] and [DLP]. The mapping $\text{dexp}|_{A}^{-1}$ is replaced by

$$\text{dcay}|_{A}^{-1} : \mathfrak{g} \to \mathfrak{g}, \quad \text{dcay}|_{A}^{-1}(B) = (I - \frac{1}{2}A)B(I + \frac{1}{2}A).$$

Alternative coordinate mappings valid for any Lie group are the so called second kind coordinates (skc) [V], and can be obtained by taking a basis $E_1, \ldots, E_d$ of $\mathfrak{g}$ as a starting point and considering the following composition of exponentials:

$$\text{skc}(A) = \exp(a_1E_1) \cdots \exp(a_dE_d), \quad A = \sum_{i=1}^{d} a_iE_i.$$ 

For further details on integration methods using these coordinates see [OM].

2.6 Applications of Lie group methods

2.6.1 Isospectral flows

Isospectral flows arise in a wide range of applications for example in certain control problems and in medical imaging. See p. 107 in [HLW].
2.7 Stiefel manifolds

Differential equations on Stiefel manifolds arise in many applications as for example statistical signal processing and neural networks, multivariate data analysis, and mechanical systems.

**Definition 2.5.** The Stiefel manifold is the set

\[
\text{St}(n,p) := \{ X \in M^{n \times p} \text{ s.t. } X^T X = I_{p \times p} \in M^{p \times p} \}
\]

where \( M^{n \times p} \) is the linear vector space of \( n \times p \) matrices and we assume \( p \leq n \), and \( I_{p \times p} \) denotes the \( p \times p \) identity matrix.

One can show that \( \text{St}(n,p) \) is a differentiable manifold.

Note that with \( n = p \), \( \text{St}(n,n) = SO(n) \), while, for \( p = 1 \), \( \text{St}(n,1) \) is the unit sphere in \( \mathbb{R}^n \). We want to show that any differential equation on the \( \text{St}(n,p) \) can be written in the form

\[
\dot{Y} = A(Y) \cdot Y, \quad Y(0) = Y_0 \in \text{St}(n,p), \tag{13}
\]

with \( A(Y) \in so(n) \). To prove this fact we start by characterizing the tangent space of \( \text{St}(n,p) \) by means of \( so(n) \).

Consider the curve \( X(t) \in \text{St}(n,p) \), by definition we have that for each \( t \),

\[
X(t) = X(t) \cdot I_{n,p},
\]

where

\[
I_{n,p} := \left[ \begin{array}{c} I_{p \times p} \\ O_{(n-p) \times p} \end{array} \right],
\]

and \( O_{(n-p) \times p} \) is the \((n-p) \times p\) zero matrix.

**Proposition 2.6.** For any smooth curve \( X(t) \in \text{St}(n,p) \) one can find a smooth curve \( X(t) = (X(t), X(t)^\perp) \in SO(n) \) such that \( X(t) = X \cdot I_{n,p} \),

where

\[
I_{n,p} := \left[ \begin{array}{c} I_{p \times p} \\ O_{(n-p) \times p} \end{array} \right],
\]

and \( O_{(n-p) \times p} \) is the \((n-p) \times p\) zero matrix.

Proof omitted

Consider \( V \in T_P \text{St}(n,p) \) with \( P \in \text{St}(n,p) \), assume \( X(t) \in \text{St}(n,p) \) is a smooth curve such that \( X(0) = P \), and \( X(0) = V \) then, from Proposition 2.6, we have

\[
V = \frac{d}{dt} (X(t), X(t)^\perp)_{I_{n,p}} \bigg|_{t=0} = (\dot{X}(t), \dot{X}(t)^\perp)_{I_{n,p}} \bigg|_{t=0} I_{n,p} = C(P)I_{n,p},
\]

with \( C(P) \in T_{(P,P^\perp)}SO(n) \), (assuming \( \dot{X}^\perp(0) = P^\perp \)). Now Proposition ?? guarantees the existence of an \( A(P) \in so(n) \) such that \( C(P) = A(P) \cdot (P, P^\perp) \), which implies that

\[
V = A(P) \cdot (P, P^\perp)I_{n,p} = A(P) \cdot P.
\]
A direct consequence of this characterization of $T_p \text{St}(n, p)$ by means of $\mathfrak{so}(n)$ is that any differential equation on $\text{St}(n, p)$ can be written in the form (13).

We can now apply Lie group integration methods to (13). The strategy is as usual to assume $Y(t) = \exp(\sigma(t))Y_0$ with $\sigma(t) \in \mathfrak{so}(n)$, and applying a classical Runge-Kutta method to the equation for $\sigma$, which also in this case is (11). Finally the RKMK methods applied to (13) assume exactly the same format as given in section 2.3, with the difference that the stage values now are $Y_i \in \text{St}(n, p)$ while, as before, the $\sigma_i$ belong to the Lie algebra.

The reader can show that the straightforward use of the commutator-free methods to (13) produces numerical approximations of the solution which belong to the Stiefel manifold.

**Example 2.7.** Equation (7) is a differential equation on the sphere of radius $\sqrt{\pi_0^T \pi_0}$ in $\mathbb{R}^3$. The use of Lie group methods on this problem guarantees that the numerical solution $\pi(t)$ has constant Euclidean norm, i.e. $\pi(t)^T \pi(t) = \pi_0^T \pi_0$, for all $t$.

**References**


