MATRIX LIE GROUPS AND CONTROL THEORY

Jimmie Lawson

Summer, 2007

Chapter 1

Introduction

Mathematical control theory is the area of application-oriented mathematics that treats the basic mathematical principles, theory, and problems underlying the analysis and design of control systems, principally those encountered in engineering. To *control* an object means to influence its behavior so as to achieve a desired goal.

One major branch of control theory is *optimization*. One assumes that a good model of the control system is available and seeks to optimize its behavior in some sense.

Another major branch treats control systems for which there is *uncertainity* about the model or its environment. The central tool is the use of *feedback* in order to correct deviations from and *stabilize* the desired behavior.

Control theory has its roots in the classical calculus of variations, but came into its own with the advent of efforts to control and regulate machinery and to develop steering systems for ships and much later for planes, rockets, and satellites. During the 1930s, researchers at Bell Telephone Laboratories developed feedback amplifiers, motiviated by the goal of assuring stability and appropriate response for electrical circuits. During the Second World War various military implementations and applications of control theory were developed. The rise of computers led to the implementation of controllers in the chemical and petroleum industries. In the 1950s control theory blossomed into a major field of study in both engineering and mathematics, and powerful techniques were developed for treating general multivariable, time-varying systems. Control theory has continued to advance with advancing technology, and has emerged in modern times as a highly developed discipline.

Lie theory, the theory of Lie groups, Lie algebras, and their applications

is a fundamental part of mathematics that touches on a broad spectrum of mathematics, including geometry (classical, differential, and algebraic), ordinary and partial differential equations, group, ring, and algebra theory, complex and harmonic analysis, number theory, and physics (classical, quantum, and relativistic). It typically relies upon an array of substantial tools such as topology, differentiable manifolds and differential geometry, covering spaces, advanced linear algebra, measure theory, and group theory to name a few. However, we will considerably simplify the approach to Lie theory by restricting our attention to the most important class of examples, namely those Lie groups that can be concretely realized as (multiplicative) groups of matrices.

Lie theory began in the late nineteenth century, primarily through the work of the Norwegian mathematician Sophus Lie, who called them "continuous groups," in contrast to the usually finite permutation groups that had been principally studied up to that point. An early major success of the theory was to provide a viewpoint for a systematic understanding of the newer geometries such as hyperbolic, elliptic, and projective, that had arisen earlier in the century. This led Felix Klein in his Erlanger Programm to propose that geometry should be understood as the study of quantities or properties left invariant under an appropriate group of geometric transformations. In the early twentieth century Lie theory was widely incorporated into modern physics, beginning with Einstein's introduction of the Lorentz transformations as a basic feature of special relativity. Since these early beginnings research in Lie theory has burgeoned and now spans a vast and enormous literature.

The essential feature of Lie theory is that one may associate with any Lie group G a Lie algebra $\mathfrak{g}$. The Lie algebra $\mathfrak{g}$ is a vector space equipped with a bilinear nonasssociative anticommutative product, called the *Lie bracket* or *commutator* and usually denoted $[\cdot, \cdot]$. The crucial and rather surprising fact is that a Lie group is almost completely determined by its Lie algebra $\mathfrak{g}$. There is also a basic bridge between the two structures given by the exponential map $\exp : \mathfrak{g} \to G$. For many purposes structure questions or problems concerning the highly complicated nonlinear structure G can be translated and reformulated via the exponential map in the Lie algebra $\mathfrak{g}$, where they often lend themselves to study via the tools of linear algebra (in short, nonlinear problems can often be linearized). This procedure is a major source of the power of Lie theory.

The two disciplines, control theory and Lie theory, come together in cer-

tain interesting classes of control problems that can be interpreted as problems on Lie groups or their coset spaces. In this case the states of the system are modeled by members of the Lie group and the controls by members of the Lie algebra, interpreted as invariant vector fields on the Lie group. There are significant advantages to interpreting problems in this framework whenever possible; these advantages include the availability of a rich arsenal of highly developed theoretical tools from Lie group and Lie algebra theory. In addition, Lie groups typically appear as matrix groups and one has available the concrete computational methods and tools of linear algebra and matrix theory.

Chapter 2

Matrix and Lie Groups

2.1 The General Linear Group

Let V be a finite dimensional vector space equipped with a complete norm $\|\cdot\|$ over the field $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. (Actually since the space V is finite dimensional, the norm must be equivalent to the usual euclidean norm, and hence complete.) Let $\mathrm{End}(V)$ denote the algebra of linear self-maps on V, and let $\mathrm{GL}(V)$ denote the *general linear group*, the group (under composition) of invertible self-maps. If $V = \mathbb{R}^n$, then $\mathrm{End}(V)$ may be identified with $M_n(\mathbb{R})$, the $n \times n$ matrices, and $\mathrm{GL}(V) = \mathrm{GL}_n(\mathbb{R})$, the matrices of nonvanishing determinant.

We endow $\mathrm{End}(V)$ with the usual *operator norm*, a complete norm defined by

$$\|A\| = \sup\{\|Av\| : \|v\| = 1\} = \sup\{\frac{\|Av\|}{\|v\|} : v \neq 0\},$$

which gives rise to the metric $d(A, B) = \|B - A\|$ on $\mathrm{End}(V)$ and, by restriction, on $\mathrm{GL}(V)$.

Exercise 2.1.1. $\|AB\| \leq \|A\| \, \|B\|$, $\|tA\| = |t| \, \|A\|$, and $\|A^n\| \leq \|A\|^n$.

Exercise 2.1.2. Show that $\mathrm{GL}(V)$ is a dense open subset of $\mathrm{End}(V)$. (Hint: The determinant function is polynomial, hence continuous, and $A - (1/n)I$ converges to A and is singular for at most finitely many values, since the spectrum of A is finite.)

Exercise 2.1.3. The multiplication and inversion on $\mathrm{GL}(V)$ are analytic,

i.e., expressible locally by power series. (Hint: the multiplication is actually polynomial, and the cofactor expansion shows that inversion is rational.)

A group G endowed with a Hausdorff topology is called a *topological group* if the multiplication map $m : G \times G \to G$ and the inversion map on G are continuous. By the preceding exercise $\mathrm{GL}(V)$ is a topological group.

2.2 The Exponential Map

We define the *exponential map* on $\mathrm{End}\, V$ by

$$\exp(A) := \sum_{n=0}^{\infty} \frac{A^n}{n!}.$$

Lemma 1. *The exponential map is absolutely convergent, hence convergent on all of* $\mathrm{End}(V)$*. Hence it defines an analytic self-map on* $\mathrm{End}(V)$*.*

Proof. $\| \sum_{n=0}^{\infty} \frac{A^n}{n!} \| \leq \sum_{n=0}^{\infty} \frac{\|A\|^n}{n!} = \exp(\|A\|)$. ∎

Absolute convergence allows us to rearrange terms and to carry out various algebraic operations and the process of differentation termwise. We henceforth allow ourselves the freedom to carry out such manipulations without the tedium of a rather standard detailed verification.

Exercise 2.2.1. (i) Show that the exponential image of a block diagonal matrix with diagonal blocks $A_1, \ldots, A_m$ is a block diagonal matrix with diagonal blocks $\exp(A_1), \ldots, \exp(A_n)$. In particular, to compute the exponential image of a diagonal matrix, simply apply the usual exponential map to the diagonal elements.
(ii) Suppose that A is similar to a diagonal matrix, $A = PDP^{-1}$. Show that $\exp(A) = P \exp(D) P^{-1}$.

Proposition 2. *If* $A, B \in \mathrm{End}\, V$ *and* $AB = BA$*, then* $\exp(A + B) = \exp A \exp B = \exp B \exp A$*.*

Proof. Computing termwise and rearranging we have

$$\begin{aligned}\exp A \exp B &= \Big(\sum_{n=0}^{\infty} \frac{A^n}{n!}\Big)\Big(\sum_{m=0}^{\infty} \frac{B^m}{m!}\Big) = \sum_{n,m=0}^{\infty} \frac{A^n B^m}{n!m!} \\ &= \sum_{k=0}^{\infty} \frac{1}{k!}\Big(\sum_{n+m=k} \frac{k!}{n!m!} A^n B^m\Big) \\ &= \sum_{k=0}^{\infty} \frac{1}{k!}\Big(\sum_{j=0}^{k} \binom{k}{j} A^j B^{k-j}\Big).\end{aligned}$$

Since A and B commute, the familiar binomial theorem yields

$$(A+B)^k = \sum_{j=1}^{k} \binom{k}{j} A^j B^{k-j},$$

and substituting into the previous expression yields the proposition. ∎

Let V, W be finite dimensional normed vector spaces and let $f : U \to W$, where U is a nonempty open subset of V. A linear map $L : V \to W$ is called the (*Frechét*) *derivative* of f at $x \in U$ if in some neighborhood of x

$$f(x+h) - f(x) = L(h) + r(h), \text{ where } \lim_{h \to 0} \frac{r(h)}{\|h\|} = 0.$$

If it exists, the derivative is unique and denoted by $df(x)$ or $f'(x)$.

Lemma 3. *The identity map on* $\mathrm{End}(V)$ *is the derivative at* 0 *of* $\exp : \mathrm{End}(V) \to \mathrm{End}(V)$*, i.e.,* $d\exp(0) = Id$*.*

Proof. For $h \in \mathrm{End}(V)$, we have

$$\exp(h) - \exp(0) = \sum_{n=0}^{\infty} \frac{h^n}{n!} - 1_V = \sum_{n=1}^{\infty} \frac{h^n}{n!} = \mathrm{Id}(h) + h^2 \sum_{n=2}^{\infty} \frac{h^{n-2}}{n!}, \text{ where}$$

$$\lim_{h \to 0} \left\| \frac{h^2 \sum_{n=0}^{\infty} \frac{h^n}{(n+2)!}}{\|h\|} \right\| \leq \lim_{h \to 0} \frac{\|h\|^2}{\|h\|} \Big(\sum_{n=2}^{\infty} \frac{1}{n!}\Big) = 0.$$

∎

Applying the Inverse Function Theorem, we have immediately from the preceding lemma

Proposition 4. *There exist neighborhoods U of 0 and V of I in $\operatorname{End} V$ such that $\exp|_U$ is a diffeomorphism onto V.*

For $A \in \operatorname{End} V$ and $r > 0$, let $B_r(A) = \{C \in \operatorname{End} V : \|C - A\| < r\}$.

Exercise 2.2.2. Show that $\exp(B_r(0)) \subseteq B_s(1_V)$ where $s = e^r - 1$. In particular for $r = \ln 2$, $\exp(B_r(0)) \subseteq B_1(1_V)$.

2.3 One-Parameter Groups

A *one-parameter subgroup* of a topological group G is a continuous homomorphism $\alpha : \mathbb{R} \to G$ from the additive group of real numbers into G.

Proposition 5. *For V a finite dimensional normed vector space and $A \in \operatorname{End} V$, the map $t \mapsto \exp(tA)$ is a one-parameter subgroup of $\operatorname{GL}(V)$. In particular, $(\exp(A))^{-1} = \exp(-A)$.*

Proof. Since sA and tA commute for any $s, t \in \mathbb{R}$, we have from Proposition 2 that $t \mapsto \exp(tA)$ is a homomorphism from the additive reals to the $\operatorname{End} V$ under multiplication. It is continuous, indeed analytic, since scalar multiplication and exp are. The last assertion follows from the homomorphism property and assures the the image lies in $\operatorname{GL}(V)$. ∎

Proposition 6. *Choose an $r < \ln 2$. Let $A \in B_r(0)$ and let $Q = \exp A$. Then $P = \exp(A/2)$ is the unique square root of Q contained in $B_1(1_V)$.*

Proof. Since $\exp(tA)$ defines a one-parameter subgroup,

$$P^2 = (\exp(A/2))^2 = \exp(A/2)\exp(A/2) = \exp(A/2 + A/2) = \exp(A) = Q.$$

Also $A \in B_r(0)$ implies $A/2 \in B_r(0)$, which implies $\exp(A/2) \in B_1(1_V)$ (Exercise 2.2.2).

Suppose two elements in $B_1(1_V)$, say $1+B$ and $1+C$ where $\|B\|, \|C\| < 1$ satisfy $(1+B)^2 = (1+C)^2$. Then expanding the squares, cancelling the 1's, and rearranging gives

$$2(B - C) = C^2 - B^2 = C(C - B)) + (C - B)B.$$

Taking norms yields

$$2\|B - C\| \leq \|C\| \, \|C - B\| + \|C - B\| \, \|B\| = (\|C\| + \|B\|)\|C - B\|.$$

This implies either $\|C\| + \|B\| \geq 2$, which is also false since each summand is less than 1, or $\|B - C\| = 0$, i.e., $B = C$. We conclude there at most one square root in $B_1(1_V)$. ∎

Lemma 7. *Consider the additive group $(\mathbb{R}, +)$ of real numbers.*

(i) If a subgroup contains a sequence of nonzero numbers $\{a_n\}$ converging to 0, then the subgroup is dense.

(ii) For one-parameter subgroups $\alpha, \beta : \mathbb{R} \to G$, the set $\{t \in \mathbb{R} : \alpha(t) = \beta(t)\}$ is a closed subgroup.

Proof. (i) Let $t \in \mathbb{R}$ and let $\varepsilon > 0$. Pick a_n such that $|a_n| < \varepsilon$. Pick an integer k such that $|t/a_n - k| < 1$ (for example, pick k to be the floor of t/a_n). Then multiplying by $|a_n|$ yields $|t - ka_n| < |a_n| < \varepsilon$. Since ka_n must be in the subgroup, its density follows.

(ii) Exercise. ∎

Exercise 2.3.1. Show that any nonzero subgroup of $(\mathbb{R}, +)$ is either dense or cyclic. (Hint: Let H be a subgroup and let $r = \inf\{t \in H : t > 0\}$. Consider the two cases $r = 0$ and $r > 0$.)

The next theorem is a converse of Proposition 5.

Theorem 8. *Every one parameter subgroup $\alpha : \mathbb{R} \to \mathrm{End}(V)$ is of the form $\alpha(t) = \exp(tA)$ form some $A \in \mathrm{End}\, V$.*

Proof. Pick $r < \ln 2$ such that exp restricted to $B_r(0)$ is a diffeomorphism onto an open subset containing $1 = 1_V$. This is possible by Proposition 4. Note that $\exp(B_r(0)) \subseteq B_1(1)$. By continuity of α, pick $0 < \varepsilon$ such that $\alpha(t) \in \exp(B_r(0))$ for all $-\varepsilon < t < \varepsilon$. Then $\alpha(1/2^k) \in \exp(B_r(0)) \subseteq B_1(1)$ for all $1/2^k < \varepsilon$.

Pick $1/2^n < \varepsilon$. Then $Q := \alpha(1/2^n) \in \exp(B_r(0))$ implies $Q = \alpha(1/2^n) = \exp(B)$ for some $B \in B_r(0)$. Set $A = 2^n B$. Then $Q = \exp((1/2^n)A)$.

Then $\alpha(1/2^{n+1})$ and $\exp(B/2)$ are both square roots of Q contained in $B_1(1)$, and hence by Proposition 6 are equal. Thus $\alpha(1/2^{n+1}) = \exp((1/2^{n+1})A)$. By induction $\alpha(1/2^{n+k}) = \exp((1/2^{n+k})A)$ for all positive integers k. By Lemma 7(ii) the two one-parameter subgroups agree on a closed subgroup, and by Lemma 7 this subgroup is also dense. Hence $\alpha(t)$ and $\exp(tA)$ agree everywhere. ∎

The preceding theorem establishes that a merely continuous one-parameter subgroup must be analytic. This is a very special case of Hilbert's fifth problem, which asked whether a locally euclidean topological group was actually an analytic manifold with an analytic multiplication. This problem was solved positively some fifty years later in the 1950's by Gleason, Montgomery, and Zippin.

Exercise 2.3.2. Show that if $\exp(tA) = \exp(tB)$ for all $t \in \mathbb{R}$, then $A = B$. (Hint: Use Proposition 4)

Remark 9. The element $A \in \operatorname{End} V$ is called the *infinitesimal generator* of the one-parameter group $t \mapsto \exp(tA)$. We conclude from the preceding theorem and remark that there is a one-to-one correspondence between one-parameter subgroups and their infinitesimal generators.

2.4 Curves in $\operatorname{End} V$

In this section we consider basic properties of differentiable curves in $\operatorname{End} V$. Let I be an open interval and let $A(\cdot) : I \to \operatorname{End} V$ be a curve. We say that A is C^r if each of the coordinate functions $A_{ij}(t)$ is C^r on $\mathbb{R}$. We define the derivative $\dot{A}(t) = \lim_{h\to 0}(1/h)(A(t+h) - A(t))$. The derivative exists iff the derivative $\dot{A}_{ij}(t)$ of each coordinate function exists, and in this case $\dot{A}(t)$ is the linear operator with coordinate functions $\frac{d}{dt}(A_{ij}(t))$.

Items (1) and (2) in the following list of basic properties for operator-valued functions are immediate consequences of the preceding characterization, and item (5) is a special case of the general chain rule.

(1) $D_t(A(t) \pm B(t)) = \dot{A}(t) \pm \dot{B}(t)$

(2) $D_t(rA(t)) = r\dot{A}(t)$.

(3) $D_t(A(t) \cdot B(t) = \dot{A}(t) \cdot B(t) + A(t) \cdot \dot{B}(t)$.

(Note: Order is important since multiplication is noncommutative.)

(4) $D_t(A^{-1}(t)) = -A^{-1}(t) \cdot \dot{A}(t) \cdot A^{-1}(t)$.

(5) If $\dot{B}(t) = A(t)$, the $D_t(B(f(t)) = f'(t)A(f(t))$.

Exercise 2.4.1. Establish properties (3) and (4). (Hints: (3) Mimic the proof of the product rule in the real case. (4) Note $A^{-1}(t)$ is differentiable if $A(t)$ is, since it is the composition with the inversion function, which is analytic, hence C^r for all r. Differentiate the equation $A(t) \cdot A^{-1}(t) = I$ and solve for $D_t(A^{-1}(t))$.)

We can also define the integral $\int_a^b A(t)\,dt$ by taking the coordinate integrals $\int_a^b A_{ij}(t)\,dt$. The following are basic properties of the integral that follow from the real case by working coordinatewise.

(6) If $B(t) = \int_{t_0}^t A(s)\,da$, then $\dot{B}(t) = A(t)$.

(7) If $\dot{B}(t) = A(t)$, the $\int_r^s A(t)\,dt = B(s) - B(r)$.

We consider curves given by power series: $A(t) = \sum_{n=0}^{\infty} t^n A_n$. Define the n^{th}-partial sum to be $S_n(t) = \sum_{k=0}^{n} t^k A_k$. The power series *converges* for some value of t if the partial sums $S_n(t)$ converge in each coordinate to some $S(t)$. This happens iff the coordinatewise real power series all converge to the coordinates of $S(t)$.

Since for an operator A, $|a_{ij}| \leq \|A\|$ for each entry a_{ij} (exercise), we have that *absolute convergence*, the convergence of $\sum_{n=1}^{\infty} |t|^n \|A_n\|$, imples the absolute convergence of each of the coordinate series, and their uniform convergence over any closed interval in the open interval of convergence of the real power series $\sum_{n=1}^{\infty} t^n \|A_n\|$. These observations justify termwise differentiation and integration in the interval of convergence of $\sum_{n=1}^{\infty} t^n \|A_n\|$.

Exercise 2.4.2. (i) Show that the power series

$$\exp(tA) = \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n$$

is absolutely convergent for all t (note that $A_n = (1/n!)A^n$ in this series).
(ii) Use termwise differentiation to show $D_t(\exp(tA)) = A\exp(tA)$.
(iii) Show that $X(t) = \exp(tA)X_0$ satisfies the differential equation on $\operatorname{End} V$ given by

$$\dot{X}(t) = AX(t), \qquad X(0) = X_0.$$

(iv) Show that the equation $\dot{x}(t) = Ax(t)$, $x(0) = x_0$ on V has solution $x(t) = \exp(tA)x_0$.

2.5 The Baker-Campbell-Hausdorff Formalism

It is a useful fact that the derivative of the multiplication map at the identity I of $\operatorname{End} V$ is the addition map.

Proposition 10. *Let $m : \operatorname{End}(V) \times \operatorname{End}(V) \to \operatorname{End}(V)$ be the multiplication map, $m(A, B) = AB$. Then the derivative at (I, I), $d_{(I,I)}m : \operatorname{End}(V) \times \operatorname{End}(V) \to \operatorname{End} V$ is given by $dm_{(I,I)}(U, V) = U + V$.*

Proof. Since the multiplication map is polynomial, continuous partials of all orders exist, and in particular the multiplication map is differentiable. By definition the value of the derivative at (I, I) evaluated at some $(U, 0) \in' \operatorname{End}(V) \times \operatorname{End}(V)$ is given by

$$dm_{(I,I)}(U, 0) = \lim_{t \to 0} \frac{m(I + tU, I) - m(I, I)}{t} = \lim_{t \to 0} \frac{tU}{t} = U.$$

∎

We have seen previously that the exponential function is a diffeomorphism from some open neighborhood B of 0 to some open neighborhood U of I. Thus there exists an analytic inverse to the exponential map, which we denote by $\log : U \to B$. Indeed if one defines

$$\log(I - A) = -\sum_{n=1}^{\infty} \frac{A^n}{n},$$

then just as for real numbers this series converges absolutely for $\|A\| < 1$. Further since $\exp(\log A) = A$ holds in the case of real numbers, it holds in the algebra of formal power series, and hence in the linear operator or matrix case. Indeed one can conclude that exp is $1 - 1$ on $B_{\ln 2}(0)$, carries it into $B_1(I)$, and has inverse given by the preceding logarithmic series, all this without appeal to the Inverse Function Theorm.

The local diffeomorphic property of the exponential function allows one to pull back the multiplication in $\mathrm{GL}(V)$ locally to a neighborhood of 0 in $\operatorname{End} V$. One chooses two points A, B in a sufficiently small neighborhood of 0, forms the product $\exp(A) \cdot \exp(B)$ and takes the log of this product:

$$A * B := \log(\exp A \cdot \exp B).$$

This *Baker-Campbell-Hausdorff multiplication* is defined on any $B_r(0)$ small enough so that $\exp(B_r(0)) \cdot \exp(B_r(0))$ is contained in the domain of the log function; such exist by the local diffeomorphism property and the continuity of multiplication. Now there is a beautiful formula called the *Baker-Campbell-Hausdorff formula* that gives $A * B$ as a power series in A and B with the higher powers being given by higher order Lie brackets or commutators, where the (first-order) *commutator* or *Lie bracket* is given by $[A, B] := AB - BA$. The Baker-Campbell-Hausdorff power series is obtained by manipulating the power series for $\log(\exp(x) \cdot \exp(y))$ in two noncommuting variables x, y in such a way that it is rewritten so that all powers are commutators of some order. To develop this whole formula would take us too far afield from our goals, but we do derive the first and second order terms, which suffice for many purposes.

Definition 11. An open ball $B_r(0)$ is called a *Baker-Campbell-Hausdorff neighborhood*, or BCH-*neighborhood* for short, if $r < 1/2$ and $\exp(B_r(0) \cdot \exp(B_r(0) \subseteq B_s(0)$ for some s, r such that exp restricted to $B_s(0)$ is a diffeomorphism onto some open neighborhood of I. By the local diffeomorphism property of the exponential map and the continuity of multiplication at I, BCH-neighborhoods always exist. We define the *Baker-Campbell-Hausdorff multiplication* on any BCH-neighborhood $B_r(0)$ by

$$A * B = \log(\exp A \cdot \exp B).$$

Note that $A * B$ exists for all $A, B \in B_r(0)$, but we can only say that $A * B \in \operatorname{End} V$, not necessarily in $B_r(0)$.

Proposition 12. *Let $B_r(0)$ be a BCH-neighborhood. Define $\mu : B_r(0) \times B_r(0) \to \operatorname{End} V$ by $\mu(A, B) = A * B$. Then*

(i) $A * B = A + B + R(A, B)$ where $\lim_{A,B \to 0} \dfrac{\|R(A,B)\|}{\|A\| + \|B\|} = 0$.

(ii) There exists $0 < s \leq r$ such that $\|A * B\| \leq 2(\|A\| + \|B\|)$ for $A, B \in B_s(0)$.

Proof. (i) We have that $\mu = \log \circ m \circ (\exp \times \exp)$, so by the chain rule, the fact that the derivatives of exp at 0 and log at I are both the identity map $Id : \operatorname{End} V \to \operatorname{End} V$ (Lemma 3 and the Inverse Function Theorem) and Proposition 10, we conclude that

$$d\mu_{(0,0)}(U, V) = Id \circ dm_{(I,I)} \circ (Id \times Id)(U, V) = U + V.$$

By definition of the derivative, we have

$$\begin{aligned} U * V &= U * V - 0 * 0 = dm_{(0,0)}(U,V) + R(U,V) \\ &= U + V + R(U,V) \quad \text{where} \lim_{(U,V)\to(0,0)} \frac{\|R(U,V)\|}{\|U\| + \|V\|} = 0. \end{aligned} \tag{2.1}$$

(Note that the second equality is just the definition of the derivative, where the norm on $\operatorname{End} V \times \operatorname{End} V$ is the sum norm.) This gives (i).

(ii) Using (i), we obtain the following string:

$$\|A * B\| \leq \|A * B - A - B\| + \|A + B\| \leq \|R(A,B)\| + \|A\| + \|B\|.$$

Now $\|R(A,B)\| \to 0$ as $A, B \to 0$, so the right-hand sum is less than or equal $2(\|A\| + \|B\|)$ on some $B_s(0) \subseteq B_r(0)$. ∎

Exercise 2.5.1. Use the fact that $0 * 0 = 0$ and imitate the proof of Proposition 10 to show directly that $dm_{(0,0)}(U,V) = U + V$.

We now derive the linear and second order terms of the Baker-Campbell-Hausdorff series.

Theorem 13. *Let $B_r(0)$ be a BCH-neighborhood. Then*

$$A * B = A + B + \frac{1}{2}[A,B] + S(A,B) \text{ where } \lim_{A,B\to 0} \frac{\|S(A,B)\|}{(\|A\| + \|B\|)^2} = 0.$$

Proof. Pick $B_s(0) \subseteq B_r(0)$ so that condition (ii) of Proposition 12 is satisfied. Setting $C = A * B$, we have directly from the definition of $A * B$ that $\exp C = \exp A \cdot \exp B$. By definition

$$\exp C = I + C + \frac{C^2}{2} + R(C), \quad \text{where} \quad R(C) = \sum_{n=3}^{\infty} \frac{C^n}{n!}. \tag{2.2}$$

For $A, B \in B_s(0)$, we have from Proposition 12 that $\|C\| = \|A * B\| \leq 2(\|A\| + \|B\|) < 1$ since $r < 1/2$. Thus we have the estimate

$$\|R(C)\| \leq \sum_{n=3}^{\infty} \frac{\|C\|^n}{n!} \leq \|C\|^3 \sum_{n=3}^{\infty} \frac{\|C\|^{n-3}}{n!} \leq \frac{1}{2}\|C\|^3. \tag{2.3}$$

Recalling the calculations in the proof of Proposition 2, we have

$$\exp A \cdot \exp B = I + A + B + \frac{1}{2}(A^2 + 2AB + B^2) + R_2(A,B), \tag{2.4}$$

$$\text{where } R_2(A,B) = \sum_{n=3}^{\infty} \frac{1}{n!} \Big(\sum_{k=0}^{n} \binom{n}{k} A^k B^{n-k} \Big).$$

We have the estimate

$$\|R_2(A,B)\| \leq \sum_{n=3}^{\infty} \frac{1}{n!} (\|A\| + \|B\|)^n \leq \frac{1}{2} (\|A\| + \|B\|)^3. \tag{2.5}$$

If in the equation $\exp C = \exp A \cdot \exp B$, we replace $\exp C$ by the right side of equation (2.2), $\exp A \cdot \exp B$ by the right side equation (2.4), and solve for C, we obtain

$$C = A + B + \frac{1}{2}(A^2 + 2AB + B^2 - C^2) + R_2(A,B) - R(C). \tag{2.6}$$

Since

$$\begin{aligned} A^2 + 2AB + B^2 - C^2 &= [A,B] + (A+B)^2 - C^2 \\ &= [A,B] + (A+B)(A+B-C) + (A+B-C)C, \end{aligned}$$

we obtain alternatively

$$C = A + B + \frac{1}{2}[A,B] + S(A,B), \tag{2.7}$$

where $S(A,B) = \frac{1}{2}((A+B)(A+B-C)+(A+B-C)C) + R_2(A,B) - R(C)$.

To complete the proof, it suffices to show that the limit as $A, B \to 0$ of each of the terms of $S(A,B)$ divided by $(\|A\| + \|B\|)^2$ is 0. First we have in $B_s(0)$

$$\begin{aligned} \frac{1}{2}\|(A+B)(A+B-C) + (A+B-C)C\| &\leq \frac{1}{2}(\|A\| + \|B\| + \|C\|)\|A+B-C\| \\ &\leq \frac{1}{2}(\|A\| + \|B\| + 2((\|A\| + \|B))\|R(A,B)\| \\ &= \frac{3}{2}(\|A\| + \|B)\|R(A,B)\|, \end{aligned}$$

where the second inequality and last equality follow by applying appropriate parts of Proposition 12. Proposition 12 also insures that

$$\lim_{A,B \to 0} \frac{3(\|A\| + \|B\|)\|R(A,B)\|}{2(\|A\| + \|B\|)^2} = 0.$$

That $\lim_{A,B\to 0} \|R_2(A,B)\|/(\|A\|+\|B\|)^2 = 0$ follows directly from equation(2.5). Finally by equation (2.3) and Proposition 12(ii)

$$\frac{\|R(C)\|}{(\|A\|+\|B\|)^2} \leq \frac{1}{2}\frac{\|C\|^3}{(\|A\|+\|B\|)^2} \leq \frac{4(\|A\|+\|B\|)^3}{(\|A\|+\|B\|)^2}$$

which goes to 0 as $A, B \to 0$. ∎

2.6 The Trotter and Commutator Formulas

In the following sections we show that one can associate with each closed subgroup of $\mathrm{GL}(V)$ a Lie subalgebra of $\operatorname{End} V$, that is, a subspace closed under Lie bracket. The exponential map carries this Lie algebra into the matrix group and using properties of the exponential map, one can frequently transfer structural questions about the Lie group to the Lie algebra, where they often can be treated using methods of linear algebra. In this section we look at some of the basic properties of the exponential map that give rise to these strong connections between a matrix group and its Lie algebra.

Theorem 14. *(Trotter Product Formula) Let $A, B \in \operatorname{End} V$ and let $\lim_n nA_n = A$, $\lim_n nB_n = B$. Then*

(i) $A + B = \lim_n n(A_n * B_n)$;

(ii) $\exp(A+B) = \lim_n(\exp A_n \exp B_n))^n = \lim_n(\exp((A/n)\exp(B/n)))^n$.

Proof. (i) Let $\varepsilon > 0$ For large n, $n\|A_n\| \leq \|A\| + \|nA_n - A\| < \|A\| + \varepsilon$, and thus $\|A_n\| \leq (1/n)(\|A\| + \varepsilon)$. It follows that $\lim_n A_n = 0$ and similarly $\lim_n B_n = 0$. By Proposition 12(i) we have

$$\lim_n n(A_n * B_n) = \lim_n nA_n + \lim_n nB_n + \lim_n nR(A_n, B_n) = A + B,$$

provided that $\lim_n nR(A_n, B_n) = 0$. But we have

$$\|nR(A_n,B_n)\| = \frac{n(\|A_n\|+\|B_n\|)\|R(A_n,B_n)\|}{\|A_n\|+\|B_n\|} \to (\|A\|+\|B\|)\cdot 0 \text{ as } n \to \infty.$$

(ii) The first equality follows directly by applying the exponential function to (i):

$$\begin{aligned}\exp(A+B) &= \exp(\lim_n n(A_n * B_n)) = \lim_n \exp(n(A_n * B_n)) \\ &= \lim_n(\exp(A_n * B_n))^n = \lim_n(\exp(A_n)\exp(B_n))^n\end{aligned}$$

where the last equality follows from the fact that exp is a local isomorphism from the BCH-multiplication to operator multiplication, and penultimate equality from the fact that $\exp(nA) = \exp(A)^n$, since exp restricted to $\mathbb{R}A$ is a one-parameter group. The second equality in part (i) of the theorem follows from the first by setting $A_n = A/n$, $B_n = B/n$. ∎

The exponential image of the Lie bracket of the commutator can be calculated from products of group commutators.

Theorem 15. *(Commutator Formula) Let $A, B \in \operatorname{End} V$ and let $\lim_n nA_n = A$, $\lim_n nB_n = B$. Then*

(i) $[A, B] = \lim_n n^2(A_n * B_n - B_n * A_n) = \lim_n n^2(A_n * B_n * (-A_n) * (-B_n))$;

(ii) $\exp[A, B] = \lim_n(\exp(A_n)\exp(B_n)(\exp A_n)^{-1}(\exp B_n)^{-1})^{n^2}$
$= \lim_n(\exp(A/n)\exp(B/n)(\exp(A/n))^{-1}(\exp(B/n))^{-1})^{n^2}$.

Proof. (i) From Theorem 13 we have for A, B in a BCH-neighborhood:

$$\begin{aligned} A * B - B * A &= \frac{1}{2}([A, B] - [B, A]) + S(A, B) - S(B, A) \\ &= [A, B] + S(A, B) - S(B, A) \end{aligned}$$

since $[A, B] = -[B, A]$. Therefore

$$\begin{aligned} \lim_n n^2(A_n * B_n - B_n * A_n) &= \lim_n n^2([A_n, B_n] + S(A_n, B_n) - S(B_n, A_n)) \\ &= \lim_n [nA_n, nB_n] + \lim_n (n^2 S(A_n, B_n) - n^2 S(B_n, A_n)), \\ &= [A, B], \end{aligned}$$

provided $\lim_n n^2 S(A_n, B_n) = \lim_n n^2 S(B_n, A_n)) = 0$. To see this, we note

$$\lim_n n^2 \|S(A_n, B_n)\| = \lim_n n^2(\|A_n\| + \|B_n)^2 \frac{\|S(A_n, B_n)\|}{(\|A_n\| + \|B_n\|)^2} = (\|A\| + \|B\|)^2 \cdot 0 = 0$$

and similarly $\lim_n n^2 \|S(B_n, A_n)\| = 0$.

To see second equality in item (i), observe first that on a BCH-neighborhood where the exponential map is injective,

$$\begin{aligned} \exp((-A) * (-B)) &= \exp(-A)\exp(-B) = (\exp A)^{-1}(\exp B)^{-1} \\ &= ((\exp B)(\exp A))^{-1} = (\exp(B * A))^{-1} = \exp(-B * A), \end{aligned}$$

which implies $(-A)*(-B) = -(B*A)$. Hence we have by Theorem 13 that

$$\begin{aligned} A*B*(-A)*(-B) - (A*B - B*A) &= (A*B)*(-(B*A)) - (A*B + (-B*A)) \\ &= \frac{1}{2}[A*B, -B*A] + S(A*B, -B*A). \end{aligned}$$

Applying this equality to the given sequences, we obtain

$$\begin{aligned} & n^2\|A_n*B_n*(-A_n)*(-B_n) - (A_n*B_n - B_n*A_n\| \\ \leq\ & \frac{n^2}{2}\|[A_n*B_n, -B_n*A_n]\| + n^2\|S(A_n*B_n, -B_n*A_n)\|. \end{aligned}$$

Now if we show that the two terms in the second expression approach 0 as $n \to \infty$, then the first expression approaches 0, and thus the two limits in (i) will be equal. We observe first that by the Trotter Product Formula

$$\lim_n n^2[A_n*B_n, -B_n*A_n] = \lim_n [nA_n*B_n, -nB_n*A_n] = [A+B, -(B+A] = 0$$

since $[C, -C] = -[C, C] = 0$ for any C. Thus the first right-hand term approaches 0. For the second

$$\begin{aligned} & n^2\|S(A_n*B_n, -B_n*A_n)\| \\ =\ & n^2(\|A_n*B_n\| + \| - B_n*A_n\|)^2 \frac{\|S(A_n*B_n, -B_n*A_n)\|}{(\|A_n*B_n\| + \| - B_n*A_n\|)^2} \\ \to\ & (\|A+B\| + \| - (B+A)\|)^2 \cdot 0 = 0 \end{aligned}$$

as $n \to \infty$.

(ii) The proof follows from an application of the exponential function to part (i), along the lines of the Trotter Product Formula. ∎

Exercise 2.6.1. Give the proof of part (ii) in the preceding theorem.

2.7 The Lie Algebra of a Matrix Group

In this section we set up the fundamental machinery of Lie theory, namely we show how to assign to each matrix group a (uniquely determined) Lie algebra and an exponential map from the Lie algebra to the matrix group that connects the two together. We begin be defining the notions and giving some examples.

By a *matrix group* we mean a closed subgroup of $\mathrm{GL}(V)$, where V is a finite dimensional vector space.

Examples 2.7.1. The following are standard and basic examples.

(1) The general linear group $\mathrm{GL}(V)$. If $V = \mathbb{R}^n$, then we write the group of $n \times n$ invertible matrices as $\mathrm{GL}_n(\mathbb{R})$.

(2) The special linear group $\{g \in \mathrm{GL}(V) : \det(g) = 1\}$.

(3) Let V be a real (resp. complex) Hilbert space equipped with an inner product $\langle \cdot, \cdot \rangle$. The orthogonal group (resp. unitary group) consists of all transformations preserving the inner product, i.e.,

$$O(V)(\text{ resp. } U(V)) = \{g \in \mathrm{GL}(V) : \forall x, y \in V,\ \langle gx, gy \rangle = \langle x, y \rangle\}.$$

If $V = \mathbb{R}^n$ (resp. $\mathbb{C}^n$) equipped with the usual inner product, then the orthogonal group O_n (resp. unitary group U_n) consists of all $g \in \mathrm{GL}(V)$ such that $g^t = g^{-1}$ (resp. $g^* = g^{-1}$).

(4) Let $V = \mathbb{R}^n \oplus \mathbb{R}^n$ equipped with the sympletic form

$$Q\Big(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}\Big) := \langle x_1, y_2 \rangle - \langle y_1, x_2 \rangle.$$

The *real sympletic group* is the subgroup of $\mathrm{GL}(V)$ preserving Q:

$$\mathrm{Sp}(V) = \mathrm{SP}_{2n}(\mathbb{R}) := \{M \in \mathrm{GL}(V) \colon \forall x, y \in V,\ Q(Mx, My) = Q(x, y)\}.$$

(5) Let $0 < m, n$ and consider the group of block upper triangular real matrices

$$U_{m,n} = \left\{ \begin{bmatrix} A & B \\ 0 & D \end{bmatrix} \in \mathrm{GL}_{m+n}(\mathbb{R}) : A \in \mathrm{GL}_m(\mathbb{R}),\ B \in M_{m,n}(\mathbb{R}),\ D \in \mathrm{GL}_n(\mathbb{R}) \right\}.$$

This is the subgroup of $\mathrm{GL}_{m+n}(\mathbb{R})$ that carries the subspace $\mathbb{R}_m \oplus \{0\}$ of $\mathbb{R}^m \oplus \mathbb{R}^n$ into itself.

Exercise 2.7.1. (i) Verify that the subgroups in (2)-(5) are closed.
(ii) Verify the alternative characterizations of elements of the subgroup in items (3) and (5).

Exercise 2.7.2. Establish the following equivalence:

- $M \in \mathrm{SP}_{2n}(\mathbb{R})$;

- $M^*JM = J$, where $J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} \in \mathrm{End}(\mathbb{R}^{2n})$;
- If M has block matrix form $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ (where all submatrices are $n \times n$), then
$$A^*C,\ B^*D \text{ are symmetric, and } A^*D - C^*B = I.$$

Definition 16. A real Lie algebra $\mathfrak{g}$ is a real vector space equipped with a binary operation
$$[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$$
satisfying the identities

(i) (Bilinearity) For all $\lambda, \mu \in \mathbb{R}$ and $X, Y, Z \in \mathfrak{g}$,
$$\begin{aligned} [\lambda X + \mu Y, Z] &= \lambda[X, Z] + \mu[Y, Z] \\ [X, \lambda Y + \mu Z] &= \lambda[X, Y] + \mu[X, Z]. \end{aligned}$$

(ii) (Skew symmetry) For all $X, Y \in \mathfrak{g}$
$$[X, Y] = -[Y, X];$$

(iii) (Jacobi identity) For all $X, Y, Z \in \mathfrak{g}$,
$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0.$$

Exercise 2.7.3. Verify that $\mathrm{End}\, V$ equipped with the Lie bracket or commutator operation $[A, B] = AB - BA$ is a Lie algebra.

It follows directly from the preceding exercise that any subspace of $\mathrm{End}\, V$ that is closed with respect to the Lie bracket operation is a Lie subalgebra.

We define a *matrix semigroup* S to be a closed multiplicative subsemigroup of $\mathrm{GL}(V)$ that contains the identity element. We define the *tangent set of* S by
$$\mathcal{L}(S) = \{A \in \mathrm{End}\, V : \exp(tA) \in S \text{ for all } t \geq 0\}.$$

We define a *wedge* in $\mathrm{End}\, V$ to be a closed subset containing $\{0\}$ that is closed under addition and scalar multiplication by nonnegative scalars.

Proposition 17. *If S is a matrix semigroup, then $\mathcal{L}(S)$ is a wedge.*

Proof. Since $I = \exp(t.0)$ for all $t \geq 0$ and $I \in S$, we conclude that $0 \in \mathcal{L}(S)$. If $A \in \mathcal{L}(S)$, then $\exp(tA) \in S$ for all $t \geq 0$, and thus $\exp(rtA) \in S$ for all $r, t \geq 0$ It follows that $rA \in \mathcal{L}(S)$ for $r \geq 0$. Finally by the Trotter Product Formula if $A, B \in \mathcal{L}(S)$, then

$$\exp(t(A+B)) = \lim_n (\exp(tA/n)\exp(tB/n))^n \in S \text{ for } t \geq 0$$

since S is a closed subsemigroup of $\mathrm{GL}(V)$. Thus $A + B \in \mathcal{L}(S)$. ∎

Theorem 18. *For a matrix group $G \subseteq \mathrm{GL}(V)$, the set*

$$\mathfrak{g} = \{A \in \operatorname{End} V : \exp(tA) \in G \text{ for all } t \in \mathbb{R}\}.$$

is a Lie algebra, called the Lie algebra of G.

Proof. As in the proof of Proposition 17, $\mathfrak{g}$ is closed under addition and scalar multiplication, i.e., a subspace of $\operatorname{End} V$. By the Commutator Formula for $A, B \in \mathfrak{g}$,

$$\exp([A,B]) = \lim_n ((\exp A/n)(\exp B/n)(\exp(-A/n)(\exp(-B/n)^{n^2} \in G$$

since G is a closed subgroup of $\mathrm{GL}(V)$. Replacing A by tA, which again is in $\mathfrak{g}$, we have $\exp(t[A,B]) = \exp([tA,B]) \in G$ for all $t \in \mathbb{R}$. Thus $[A,B] \in \mathfrak{g}$. ∎

Exercise 2.7.4. Show for a matrix group G (which is a matrix semigroup, in particular) that $\mathfrak{g} = \mathcal{L}(G)$.

Lemma 19. *Suppose that G is a matrix group, $\{A_n\}$ is a sequence in $\operatorname{End} V$ such that $A_n \to 0$ and $\exp(A_n) \in G$ for all n. If $s_n A_n$ has a cluster point for some sequence of real numbers s_n, then the cluster point belongs to $\mathfrak{g}$.*

Proof. Let B be a cluster point of $s_n A_n$. By passing to an appropriate subsequence, we may assume without loss of generality that $s_n A_n$ converges to B. Let $t \in \mathbb{R}$ and for each n pick an integer m_n such that $|m_n - ts_n| < 1$. Then

$$\begin{aligned} \|m_n A_n - tB\| &= \|(m_n - ts_n)A_n + t(s_n A - B)\| \\ &\leq |m_n - ts_n|\,\|A_n\| + |t|\,\|s_n A_n - B\| \\ &\leq \|A_n\| + |t|\,\|s_n A_n - B\| \to 0, \end{aligned}$$

which implies $m_n A_n \to tB$. Since $\exp(m_n A_n) = (\exp A_n)^{m_n} \in G$ for each n and G is closed, we conclude that the limit of this sequence $\exp(tB)$ is in G. Since t was arbitrary, we see that $B \in \mathfrak{g}$. ∎

We come now to a crucial and central result.

Theorem 20. *Let $G \subseteq \mathrm{GL}(V)$ be a matrix group. Then all sufficiently small open neighborhoods of 0 in $\mathfrak{g}$ map homeomorphically onto open neighborhoods of I in G.*

Proof. Let $B_r(0)$ be a BCH-neighborhood around 0 in $\operatorname{End} V$, which maps homeomorphically under exp to an open neighborhood $\exp(B_r(0))$ of I in $\mathrm{GL}(V)$ with inverse log. Assume that $\exp(B_r(0) \cap \mathfrak{g})$ does not contain a neighborhood of I in G. Then there exists a sequence g_n contained in G but missing $\exp(B_r(0) \cap \mathfrak{g})$ that converges to I. Since $\exp(B_r(0))$ is an open neighborhood of I, we may assume without loss of generality that the sequence is contained in this open set. Hence $A_n = \log g_n$ is defined for each n, and $A_n \to 0$. Note that $A_n \in B_r(0)$, but $A_n \notin \mathfrak{g}$, for each n, since otherwise $\exp(A_n) = g_n \in \exp(\mathfrak{g} \cap B_r(0))$.

Let W be a complementary subspace to $\mathfrak{g}$ in $\operatorname{End} V$ and consider the restriction of the BCH-multiplication $\mu(A, B) = A * B$ to $(\mathfrak{g} \cap B_r(0)) \times (W \cap B_r(0))$. By the proof of Proposition 12, the derivative $d\mu_{(0,0)}$ of μ at $(0, 0)$ is addition, and so the derivative of the restriction of μ to $(\mathfrak{g} \cap B_r(0)) \times (W \cap B_r(0))$ is the addition map $+ : \mathfrak{g} \times W \to \operatorname{End} V$. Since $\mathfrak{g}$ and W are complementary subspaces, this map is an isomorphism of vector spaces. Thus by the Inverse Function Theorem there exists an open ball $B_s(0)$, $0 < s \leq r$, such that μ restricted to $(\mathfrak{g} \cap B_s(0)) \times (W \cap B_s(0))$ is a diffeomorphism onto an open neighborhood Q of 0 in $\operatorname{End} V$. Since $A_n \in Q$ for large n, we have $A_n = B_n * C_n$ (uniquely) for $B_n \in (\mathfrak{g} \cap B_s(0))$ and $C_n \in (W \cap B_s(0)$. Since the restriction of μ is a homeomorphism and $0 * 0 = 0$, we have $(B_n, C_n) \to (0, 0)$, i.e., $B_n \to 0$ and $C_n \to 0$.

By compactness of the unit sphere in $\operatorname{End} V$, we have that $C_n / \|C_n\|$ clusters to some $C \in W$ with $\|C\| = 1$. Furthermore,

$$g_n = \exp(A_n) = \exp(B_n * C_n) = \exp(B_n) \exp(C_n)$$

so that $\exp(C_n) = (\exp B_n)^{-1} g_n \in G$. It follows from Lemma 19 that $C \in \mathfrak{g}$. But this is impossible since $\mathfrak{g} \cap W = \{0\}$ and $C \neq 0$. We conclude that $\exp(B_r(0) \cap \mathfrak{g})$ does contain some neighborhood N of I in G.

Pick any open neighborhood $U \subset (B_r(0) \cap \mathfrak{g})$ of 0 in $\mathfrak{g}$ such that $\exp(U) \subseteq N$. Then $\exp U$ is open in $\exp(B_r(0) \cap \mathfrak{g})$ (since exp restricted to $B_r(0)$ is a homeomorphism), hence is open in N, and thus is open in G, being an open subset of an open set. ∎

Although we treat matrix groups from the viewpoint of elementary differential geometry in Chapter 5, we sketch here how that theory of matrix groups develops from what we have already done in that direction. Recall that a manifold is a topological space M, which we will assume to be metrizable, that has a covering of open sets each of which is homeomorphic to an open subsets of euclidean space. Any family of such homeomorphisms from any open cover of M is called an atlas, and the members of the atlas are called charts. The preceding theorem allows us to introduce charts on a matrix group G in a very natural way. Let U be an open set around 0 in $\mathfrak{g}$ contained in a BCH-neighborhood such that $W = \exp U$ is an open neighborhood of I in G. Let $\lambda_g : G \to G$ be the left translation map, i.e., $\lambda_g(h) = gh$. We define an atlas of charts on G by taking all open sets $g^{-1}N$, where N is an open subset of G such that $I \in N \subseteq W$ and defining the chart to be $\log \circ \lambda_g : g^{-1}N \to \mathfrak{g}$ (to view these as euclidean charts, we identify $\mathfrak{g}$ with some $\mathbb{R}^n$ via identifying some basis of $\mathfrak{g}$ with the standard basis of $\mathbb{R}^n$). One can check directly using the fact that multiplication of matrices is polynomial that for two such charts ϕ and ψ, the composition $\phi \circ \psi^{-1}$, where defined, is smooth, indeed analytic. This gives rise to a differentiable structure on G, making it a smooth (analytic) manifold. The multiplication and inversion on G, when appropriately composed with charts are analytic functions, and thus one obtains an analytic group, a group on an analytic manifold with analytic group operations. This is the unique analytic structure on the group making it a smooth manifold so that the exponential map is also smooth.

2.8 The Lie Algebra Functor

We consider the category of matrix groups to be the category with objects matrix groups and morphisms continuous (group) homomorphisms and the category of Lie algebras with objects subalgebras of some $\operatorname{End} V$ and morphisms linear maps that preserve the Lie bracket., The next result shows that the assignment to a matrix group of its Lie algebra is functorial.

Proposition 21. *Let $\alpha : G \to H$ be a continuous homomorphism between matrix groups. Then there exists a unique Lie algebra homomorphism*

$d\alpha : \mathfrak{g} \to \mathfrak{h}$ *such that the following diagram commutes:*

$$\begin{array}{ccc} G & \xrightarrow{\alpha} & H \\ {\scriptstyle \exp}\uparrow & & \uparrow{\scriptstyle \exp} \\ \mathfrak{g} & \xrightarrow{d\alpha} & \mathfrak{h}. \end{array}$$

Proof. Let $A \in \mathfrak{g}$. Then the map $\beta(t) := \alpha(\exp(tA))$ is a one-parameter subgroup of H. Hence it has a unique infinitesimal generator $\tilde{A} \in \mathfrak{h}$. Define $d\alpha(A) = \tilde{A}$. We show that $d\alpha$ is a Lie algebra homomorphism. For $r \in \mathbb{R}$,

$$\alpha(\exp(trA)) = \exp(tr\tilde{A}),$$

so the infinitesimal generator for the left-hand side is $r\tilde{A}$. This shows that $d\alpha(rA) = r\tilde{A} = rd\alpha(A)$, so $d\alpha$ is homogeneous.

Let $A, B \in G$. Then

$$\begin{aligned} (\alpha \circ \exp)(t(A+B)) &= (\alpha \circ \exp)(tA + tB) = \alpha(\lim_n (\exp(tA/n)\exp(tB/n))^n) \\ &= \lim_n (\alpha(\exp(tA/n))\alpha(\exp(tB/n)))^n \\ &= \lim_n (\exp(t\tilde{A}/n)\exp(t\tilde{B}/n))^n \\ &= \exp(t\tilde{A} + t\tilde{B}) = \exp(t(\tilde{A} + \tilde{B})). \end{aligned}$$

This shows that $d\alpha(A+B) = \tilde{A} + \tilde{B} = d\alpha(A) + d\alpha(B)$, and thus $d\alpha$ is linear. In an analogous way using the commutator, one shows that $d\alpha$ preserves the commutator.

If $d\alpha(A) = \tilde{A}$, then by definition for all t, $\alpha(\exp(tA)) = \exp(t\tilde{A})$. For $t = 1$, $\alpha(\exp A) = \exp(\tilde{A}) = \exp(d\alpha(A))$. Thus $\alpha \circ \exp = d\alpha \circ \exp$. This shows the square commutes. If $\gamma : \mathfrak{g} \to \mathfrak{h}$ is another Lie algebra homomorphism that also makes the square commute, then for $A \in \mathfrak{g}$ and all $t \in \mathbb{R}$,

$$\exp(td\alpha(A)) = \exp(d\alpha(tA)) = \alpha(\exp(tA)) = \exp(\gamma(tA)) = \exp(t\gamma(A)).$$

The uniqueness of the infinitesimal generator implies $d\alpha(A) = \gamma(A)$, and hence $d\alpha = \gamma$. ∎

Exercise 2.8.1. Show that $d\alpha$ preserves the commutator.

Exercise 2.8.2. Let $\alpha : G \to H$ be a continuous homomorphism of matrix groups. Then the kernel K of α is a matrix group with Lie algebra the kernel of $d\alpha$.

2.9 Computing Lie Algebras

In this section we consider some tools for computing the Lie algebra of a matrix group, or more generally a closed subsemigroup of a matrix group. We begin with a general technique.

Proposition 22. *Let $\beta(\cdot,\cdot)$ be a continuous bilinear form on V and set*

$$G = \{g \in \mathrm{GL}(V) : \forall x, y \in V,\ \beta(gx, gy) = \beta(x, y)\}.$$

Then

$$\mathfrak{g} = \{A \in \mathrm{End}\, V : \forall x, y \in V,\ \beta(Ax, y) + \beta(x, Ay) = 0\}.$$

Proof. If $A \in \mathfrak{g}$, then $\beta(\exp(tA)x, \exp(tA)y) = \beta(x, y)$ for all $x, y \in V$. Differentiating the equation with respect to t by the product rule (which always holds for continuous bilinear forms), we obtain

$$\beta(A\exp(tA)x, \exp(tA)y) + \beta(\exp(tA)x, A\exp(tA)y) = 0.$$

Evaluating at $t = 0$ yields $\beta(Ax, y) + \beta(x, Ay) = 0$.

Conversely suppose for all $x, y \in V$, $\beta(Ax, y) + \beta(x, Ay) = 0$. Then from the computation of the preceding paragraph the derivative of

$$f(t) := \beta(\exp(tA)x, \exp(tA)y)$$

is $f'(t) = 0$. Thus f is a constant function with the value $\beta(x, y)$ at 0. It follows that $\exp(tA) \in G$ for all t, i.e., $A \in \mathfrak{g}$. ∎

Exercise 2.9.1. Apply the preceding proposition to show that the Lie algebra of the orthogonal group $O_n(\mathbb{R})$ (resp. the unitary group $U_n(\mathbb{C})$) is the Lie algebra of $n \times n$ real (resp. complex) skew symmetric matrices .

Exercise 2.9.2. (i) Use the Jordan decomposition to show for any $A \in M_n(\mathbb{C})$, $\exp(\mathrm{tr}\, A) = \det(\exp A)$.
(ii) Use (i) and Exercise 2.8.2 to show that the Lie algebra of the group $SL_n(\mathbb{C})$ of complex matrices of determinant one is the Lie algebra of matrices of trace 0. (Hint: the determinant mapping is a continuous homomorphism from $\mathrm{GL}_n(\mathbb{C})$ to the multiplicative group of non-zero complex numbers.)
(iii) Observe that $\mathcal{L}(G \cap H) = \mathcal{L}(G) \cap \mathcal{L}(H)$. What is the Lie algebra of $\mathrm{SU}_n(\mathbb{C})$, the group of unitary matrices of determinant one?

Exercise 2.9.3. Let $V = \mathbb{R}^n \oplus \mathbb{R}^n$ equipped with the canonical sympletic form

$$Q\Big(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}\Big) := \langle x_1, y_2\rangle - \langle y_1, x_2\rangle.$$

The Lie algebra of $\mathrm{Sp}(V)$ is given by

$$\mathfrak{sp}(V) = \left\{ \begin{bmatrix} A & B \\ C & D \end{bmatrix} : D = -A^*,\, B = B^*,\, C = C^* \right\}.$$

(Hint: If $(\exp tA)^* J(\exp tA) = J$ for all t, differentiate and evaluate at $t = 0$ to obtain $A^*J + JA = 0$. Multiply this out to get the preceding conditions. Conversely any block matrix satisfying the conditions can be written as

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & -A^* \end{bmatrix} + \begin{bmatrix} 0 & B \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ C & 0 \end{bmatrix}.$$

Show directly that each of the summands is in $\mathfrak{sp}(V)$ and use the fact that $\mathfrak{sp}(V)$ is a subspace.)

We introduce another general technique, this time one that applies to semigroups and groups.

Proposition 23. *Let W be a closed convex cone in the vector space $\mathrm{End}\, V$ that is also closed under multiplication. Then $S := (I + W) \cap \mathrm{GL}(V)$ is a closed subsemigroup of $\mathrm{GL}(V)$ and $\mathcal{L}(S) = W$.*

Proof. Let $X, Y \in W$. Then $(I + X)(I + Y) = I + X + Y + XY \in I + W$ since W is closed under multiplication and addition. Thus $I + W$ is a closed subsemigroup, and thus its intersection with $\mathrm{GL}(V)$ is a subsemigroup closed in $\mathrm{GL}(V)$.

Let $A \in W$. Then for $t \geq 0$, $\exp(tA) = I + \sum_{n=1}^{\infty} t^n A^n/n!$ has all finite partial sums in $I + W$ since W is closed under multiplication, addition, and scalar multiplication. Since the whole sum is the limit, it follows that $\exp(tA)$ is in the closed set $I + W$, and since the exponential image is invertible, it is in S. Thus $A \in \mathcal{L}(S)$.

Conversely assume that $\exp(tA) \in S$ for all $t \geq 0$. Then

$$A = \frac{d}{dt}\Big|_{t=0} \exp(tA) = \lim_{t \to 0^+} \frac{\exp(tA) - I}{t} \in W,$$

where the last assertion follows from the fact that $\exp(tA) \in I + W$ for $t > 0$, and hence $\exp(tA) - I$ and therefore $(1/t)(\exp(tA) - I)$ are in W. Since W is closed the limit is also in W. ∎

Exercise 2.9.4. Use Proposition 23 to show the following in $\mathrm{GL}_n(\mathbb{R})$ or $\mathrm{GL}_n(\mathbb{C})$.

(i) The group of unipotent (diagonal entries all 1) upper triangular matrices has Lie algebra the set of strictly upper triangular matrices.

(ii) The group of invertible upper triangular matrices has Lie algebra the set of all upper triangular matrices.

(iii) The group of stochastic matrices (invertible matrices with all row sums 1) has Lie algebra the set of matrices with all row sums 0.

(iv) The semigroup S of all invertible matrices with all entries nonnegative has as its Lie wedge the set of matrices whose nondiagonal entries are nonnegative.

Chapter 3

Dynamics and Control on Matrix Groups

3.1 Time-Varying Systems

A *time-varying linear dynamical system* on $\mathbb{R}^n$ is one determined by a differential equation on $\mathbb{R}^n$ of the form

$$\dot{x}(t) = A(t)x(t), \quad x(t_0) = x_0, \tag{3.1}$$

where $A(t)$ is a measurable function (i.e., all coordinate functions are measurable) in $M_n(\mathbb{R})$ and $\|A(t)\|$ is bounded (essentially bounded suffices) on any finite subinterval of its domain. (One can also consider analogous systems on $\mathbb{C}^n$ with $A(t) \in M_n(\mathbb{C})$.) A *solution* of (3.1) is an absolutely continuous function $x(\cdot)$ from the domain of $A(\cdot)$ into $\mathbb{R}^n$ that satisfies the differential equation a.e. and the initial condition $x(t_0) = x_0$.

Remark 24. It is a standard result that follows readily from the basic results on existence and uniqueness of differential equations that the differential equation (3.1) has a unique global solution on any interval I (finite or infinite) containing t_0 on which it is defined. (See, for example, Appendix C.4 of *Mathematical Control Theory* by E. Sontag.) Actually, if $A(t)$ is not defined on all of $\mathbb{R}$, then one can extend it to all of $\mathbb{R}$ by defining it to be 0 outside its given domain of definition and obtain a global solution, so we typically consider equations defined for all t.

Exercise 3.1.1. (i) Consider the vector space of absolutely continuous functions from $\mathbb{R}$ into $\mathbb{R}^n$ with the pointwise operations of addition and scalar multiplication. Show for fixed t_0 that the assignment to $x \in \mathbb{R}^n$ the solution of (3.1) for initial condition $x(t_0) = x$ is an injective linear map from $\mathbb{R}^n$ to the vector space of absolutely continuous functions.
(ii) Show that the space of absolutely continuous functions that satisfy (3.1) with no initial condition specified is n-dimensional. (Hint: Any solution is global and hence must assume some value at time t_0.)

A convenient way to study all solutions of a time-varying linear system for all initial values simultaneously is to introduce the *fundamental differential equation*

$$\dot{X}(t) = A(t)X(t), \quad X(s) = I, \tag{3.2}$$

where $X(t) \in M_n(\mathbb{R})$ for each $t \in \mathbb{R}$. Note the fundamental differential equation arises from the given time-varying linear one.

Remark 25. (i) Note that $X(\cdot)$ satisfies (3.2) if and only if each k^{th}-column of $X(\cdot)$ satisfies (3.1) with initial condition $x(s)$ equal to the k^{th}-unit vector. It follows that (3.2) has a global solution (consisting of the matrix of column solutions). (ii) It follows from Exercise 3.1.1 that the columns of $X(\cdot)$ form a basis for the set of solutions of (3.1).

We denote the solution $X(t)$ of (3.2) with initial condition $X(s) = I$ by $\Phi(t, s)$, which is defined for all $t, s \in \mathbb{R}$. For a given system of the form (3.1), $\Phi(\cdot, \cdot)$ (or just Φ) is called the *fundamental solution* or the *transition matrix*. By definition it satisfies the defining partial differential equation

$$\frac{\partial \Phi(t, s)}{\partial t} = A(t)\Phi(t, s) \quad \Phi(s, s) = I. \tag{3.3}$$

Exercise 3.1.2. Establish the following basic properties of the fundamental solution.

(i) $\Phi(t, t) = I$.

(ii) $\Phi(t, s) = \Phi(t, r)\Phi(r, s)$. (Hint: Use uniqueness of solutions.)

(iii) $\Phi(\tau, \sigma)^{-1} = \Phi(\sigma, \tau)$. (Hint: Use (ii) and (i).)

(iv) $\frac{\partial \Phi(t,s)}{\partial s} = -\Phi(t, s)A(s)$.

Exercise 3.1.3. (i) Show that $x(t) := \Phi(t, t_0)x_0$ is the (unique) solution of equation (3.1) and $X(t) = \Phi(t, t_0)X_0$, where $X_0 \in M_n(\mathbb{R})$, is the unique solution of

$$\dot{X}(t) = A(t)X(t), \quad X(t_0) = X_0. \tag{3.4}$$

(ii) Show that the equation (3.4) is *right invariant* in the sense that if $X(t)$ is a solution for initial condition $X(t_0) = X_0$, then $Y(t) = X(t)C$ is a solution for (3.4) with initial condition $Y(t_0) = X_0C$, where $C \in M_n(\mathbb{R})$.

A special case of the preceding is the case that $A(t) = A$ is a constant map. By the chain rule

$$\frac{d}{dt}\exp((t-s)A) = A\exp((t-s)A),$$

so $X(t) = \exp((t-s)A)$ is the (unique) solution of

$$\dot{X}(t) = AX(t), \quad X(s) = I.$$

Thus we have the following

Proposition 26. *The linear differential equation $\dot{X}(t) = AX(t),\ X(s) = I$ has unique solution*

$$\Phi(t, s) = \exp((t-s)A)$$

Exercise 3.1.4. (i) Let $a < b < c$ and let $A : [a, c] \to M_n(\mathbb{R})$ be defined by $A(t) = A_1$ for $a \le t \le b$ and $A(t) = A_2$ for $b < t \le c$. Show that $\Phi(t, a) = \exp((t-a)A_1)$ for $a \le t \le b$ and $\Phi(t, a) = \exp((t-b)A_2)\exp((b-a)A_1)$ for $b < t \le c$.
(ii) Let $0 = t_0 < t_1 \ldots < t_n = T$. Suppose that $A(t) = A_i$ on $(t_{i-1}, t_i]$ for $i = 1, \ldots, n$. Generalize part (i) to determine $\Phi(t, 0)$ on $[0, T]$.

More generally, one can consider nonhomogeneous equations on on $\mathbb{R}^n$. These have a solution given in terms of the fundamental solution by the *variation of parameters formula*.

Proposition 27. *The nonhomogeneous equation*

$$\dot{x}(t) = A(t)x(t) + f(t), \quad x(t_0) = x_0, \tag{3.5}$$

on some interval J, where $f : J \to \mathbb{R}^n$ is measurable and locally bounded, has solution

$$x(t) = \Phi(t, t_0)x_0 + \int_{t_0}^{t} \Phi(t, s)f(s)ds, \tag{3.6}$$

Proof. Set $z(t) = \int_{t_0}^{t} \Phi(t_0, \tau) f(\tau) d\tau + x_0$. Differentiating with respect to t yields $\dot{z}(t) = \Phi(t_0, t) f(t)$. Define $x(t) = \Phi(t, t_0) z(t)$. Then

$$\dot{x}(t) = A(t)\Phi(t, t_0)z(t) + \Phi(t, t_0)\Phi(t_0, t)f(t) = A(t)x(t) + f(t),$$

where the second equality follows from the basic properties of Φ. ∎

3.2 Control Systems on $\mathrm{GL}_n(\mathbb{R})$

In this section we introduce some basic ideas of control theory in the specific context of control systems on the group $\mathrm{GL}_n(\mathbb{R})$. The material carries through with minor modification for $\mathrm{GL}_n(\mathbb{C})$ or even $\mathrm{GL}(V)$, but for convenience and brevity we restrict to the real case.

Let Ω be some nonempty subset of $M_n(\mathbb{R})$, called the *controls*. For any interval J of real numbers we consider the set $\mathcal{U}(J, \Omega)$ of locally bounded measurable functions, called *control functions*, from J into Ω. Each member $U \in \mathcal{U}(J, \Omega)$ determines a corresponding time varying linear differential equation, called the *fundamental control equation*,

$$\dot{X}(t) = U(t)X(t)$$

of the type that we studied in the previous section. The function U is called a *control* or *steering function* and the resulting solutions *trajectories*. By the results of the previous section we have that the solution of

$$\dot{X}(t) = U(t)X(t), \quad X(t_0) = X_0$$

is given by $X(t) = \Phi_U(t, t_0)X_0$, where Φ_U is the the fundamental solution for the fundamental equation associated to the coefficient function U. The control set Ω and the differential equation $\dot{X} = UX$ determine a *control system*.

Given a control system arising from Ω and $A, B \in \mathrm{GL}(V)$, we say that B is *reachable* or *attainable* from A if there exists an interval $[a, b]$ such that a solution of the fundamental control equation X satisfies $X(a) = A$ and $X(b) = B$ for some control function U. The set of all points reachable from A (including A itself) Is called the *reachable set* fo A, and denoted $\mathcal{R}_A$. If we put the focus on B instead of A, then instead of saying that B is reachable from A, we say that A is *controllable* to B or can be *steered* to B. The *controllability set* of B consists of all A that can be steered to B.

Exercise 3.2.1. Show that if B is reachable from A, then there exists a control function $\tilde{U} : [0, T] \to \Omega$ for some $T > 0$ such that the corresponding solution $\tilde{X}$ satisfies $X(0) = A$ and $X(T) = B$. (Hint: Consider $\tilde{U}(t) = U(t + a)$.)

Remark 28. In light of the previous exercise, we can without loss of generality define the reachable set $\mathcal{R}_A$ to be all B that arise as values $X(t)$ for $t > 0$ in solutions to the fundamental control equation for initial condtion $X(0) = A$. We henceforth adopt this approach.

Let $U_i : [0, T_i] \to \Omega$ be control functions for $i = 1, 2$. Define the *concatenation* $U_2 * U_1 : [0, T_1 + T_2] \to \Omega$ by

$$U_2 * U_1(t) = \begin{cases} U_1(t), & \text{if } 0 \leq t \leq T_1; \\ U_2(t - T_1), & \text{if } T_1 < t \leq T_1 + T_2. \end{cases} \tag{3.7}$$

We consider some elementary properties of reachable sets.

Proposition 29. *Let Ω determine a control system on $\mathrm{GL}_n(\mathbb{R})$.*

(i) If B is reachable from A, then BC is reachable from AC.

(ii) $\mathcal{R}_{AC} = \mathcal{R}_A \cdot C$.

(iii) If B is reachable from A and C is reachable from B, then C is reachable from A.

(iv) The reachable set from I is a subsemigroup S_Ω of $\mathrm{GL}_n(\mathbb{R})$.

(v) For any $A \in \mathrm{GL}_n(\mathbb{R})$, $\mathcal{R}_A = S_\Omega \cdot A$.

(vi) $\Omega \subseteq \mathcal{L}(\overline{S}_\Omega)$.

Proof. (i) Let $U : [0, T] \to \Omega$ be a control function such that $X_U(0) = A$ and $X_U(T) = B$. By the right-invariance of the system (Exercise 3.1.3(ii)), $\tilde{X}(t) = X_U(t) \cdot C$ is a solution of the fundamental control equation, which clearly starts at AC and ends at BC.

(ii) It is immediate from part (i) that $\mathcal{R}_{AC} \supseteq \mathcal{R}_A \cdot C$. But then we have

$$\mathcal{R}_{AC} = \mathcal{R}_{AC} C^{-1} C \subseteq \mathcal{R}_{ACC^{-1}} \cdot C = \mathcal{R}_A \cdot C.$$

(iii) Let $U_i : [0, T_i] \to \Omega$ be the control functions for X_i for $i = 1, 2$, where $X_1(0) = A$, $X_1(T_1) = B$, $X_2(0) = B$, and $X_2(T_2) = C$. Then X_2X_1 defined by $X_2 * X_1(t) = X_1(t)$ for $0 \leq t \leq T_1$ and $X_2 * X_1(t) = X_2(t - T_!)$ for $T_1 \leq t \leq T_2$ is an absolutely continuous function that satisfies the fundamental control equation on $[0, T_1 + T_2]$ for the control function $U_2 * U_1$ and also satisfies $X_2 * X_1(0) = A$, $X_2 * X_1(T_1 + T_2) = C$.

(iv) See the following exercise.

(v) This follows immediately from (iv) and (ii).

(vi) Since $X(t) = \exp(tA)$ is a solution of the fundamental control equation with $X(0) = I$, we conclude that $\exp(tA) \in \mathcal{L}(\overline{S}_\Omega)$ for all $t \geq 0$. ∎

Exercise 3.2.2. Consider the set $\mathcal{S}$ of all control functions $U : [0, T_U] \to \Omega$ equipped with the operation of concatenation.

(i) Show that $\mathcal{S}$ is a semigroup (sometimes called the *Myhill semigroup*).

(ii) Argue that $U_2 * U_1$ with initial condition $X(0) = I$ has solution $X(t) = X_1(t)$ for $0 \leq t \leq T_!$ and $X(t) = X_2(t - T_1) \cdot X_1(T_1)$ for $T_1 \leq t \leq T_1 + T_2$.

(iii) Define a map $\omega : \mathcal{S} \to \mathrm{GL}_n(\mathbb{R})$ by $\omega(U) = X(T_U)$, where X is the solution of the fundamental control equation for $U : [0, T_U] \to \Omega$ with initial condition $X(0) = I$. (Equivalently, $\omega(U) = \Phi_U(T_U, 0)$.) Show that ω is a homomorphism from the Myhill semigroup into $\mathrm{GL}_n(\mathbb{R})$, and hence that the image $\mathcal{R}_I$ is a subsemigroup.

We consider the Hilbert space $\mathcal{H}(T)$ of coordinatewise square integrable functions from $[0, T]$ in $M_n(\mathbb{R})$, where the inner product is the sum of the usual $L_2[0, T]$ inner products in each coordinate. We endow this Hilbert space with its weak topology. Let Ω be a bounded set in $M_n(\mathbb{R})$, and let $\mathcal{U}(\Omega, T)$ be all measurable functions from $[0, T]$ into Ω. Since Ω is bounded, these control functions are all in $\mathcal{H}(T)$. Define $\pi(A, t, U)$ for $A \in \mathrm{GL}_n(\mathbb{R})$, $0 \leq t \leq T$, and $U \in \mathcal{U}(\Omega, T)$ to be $X(T)$, where X is the solution of the fundamental control equation with initial condition $X(0) = A$. The following is a basic and standard fact about continuous dependence of solutions on controls and initial conditions.

Proposition 30. *The mapping* $(A, t, U) \mapsto \pi(A, t, U)$ *from* $\mathrm{GL}_n(\mathbb{R}) \times [0, T] \times \mathcal{U}(\Omega, T)$ *into* $\mathrm{GL}_n(\mathbb{R})$ *is continuous for each* $T > 0$*, where* $\mathcal{U}(\Omega, T)$ *is given the topology of weak convergence.*

In the preceding, recall that a sequence U_n converges weakly to U, written $U_n \xrightarrow{w} U$, if $\langle U_n, V\rangle \to \langle U, V\rangle$ for all $V \in \mathcal{H}(T)$, where $\langle \cdot, \cdot \rangle$ is the inner product on the HIlbert space $\mathcal{H}(T)$.

A control $U : [0, T] \to \Omega$ is said to be *piecewise constant* if there exists a partition of $[0, T]$, $0 = t_0 < t_1 < \ldots < t_n = T$, such that U is constant on (t_{i-1}, t_i) for $1 \leq i \leq n$. The following is another useful basic fact from measure theory.

Proposition 31. *The piecewise constant functions with values in Ω are weakly dense in the set of all control functions $\mathcal{U}(\Omega, T)$.*

Exercise 3.2.3. For square integrable functions $f, g : [0, T] \to M_n(\mathbb{R})$, the inner product in $\mathcal{H}(T)$ is given by

$$\langle f, g\rangle = \sum_{1\leq i,j\leq n} \int_0^T f_{ij}(t)g_{ij}(t)dt.$$

Show that this can be written alternatively as

$$\langle f, g\rangle = \int_0^T \mathrm{tr}(f(t)g(t)^*)dt,$$

where the integrand is the trace of the matrix product $f(t)g(t)^*$.

3.3 Control Systems on Matrix Groups

In this section we continue to consider control systems like those in the previous section given by the fundamental control equation and a set $\Omega \subseteq M_n(\mathbb{R})$.

We associate with any control set Ω a matrix group $G(\Omega)$ and a closed subsemigroup $S(\Omega)$.

Definition 32. Let Ω be a control set, a nonempty subset of $M_n(\mathbb{R})$. We define $G(\Omega)$ to be the smallest closed subgroup of $\mathrm{GL}_n(\mathbb{R})$ containing $\{\exp(tA) : t \in \mathbb{R},\ A \in \Omega\}$, that is,

$$G(\Omega) := \overline{\langle\{\exp(tA : t \in \mathbb{R},\ A \in \Omega\}\rangle_{gp}} = \overline{\langle \exp(\mathbb{R}\Omega)\rangle_{gp}},$$

where $\langle Q\rangle_{gp}$ denotes the subgroup generated by Q, which consists of all finite products of members of $Q \cup Q^{-1}$. Similarly we define $S(\Omega)$ to be the smallest

closed subsemigroup of $\mathrm{GL}_n(\mathbb{R})$ containing all $\{\exp(tA) : t \geq 0,\ A \in \Omega\}$, that is,

$$S(\Omega) := \overline{\langle \{\exp(tA) : t \geq 0,\, A \in \Omega\}\rangle}_{sgp} = \overline{\langle \exp(\mathbb{R}^+\Omega)\rangle}_{sgp},$$

where the semigroup generated by a set consists of all finite products of members of the set. We call $G(\Omega)$ the matrix semigroup *infinitesimally generated* by Ω and $S(\Omega)$ the closed semigroup *infinitesimally generated* by Ω.

Lemma 33. *For a control set Ω, let $S_{pc}(\Omega)$ denote all points reachable from the identity with piecewise constant controls. Then*

$$S_{pc}(\Omega) = \{\Phi_U(T,0) : U \text{ is piecewise constant}\} = \langle \exp(\mathbb{R}^+\Omega)\rangle_{sgp}.$$

Proof. Let B be reachable from I by a piecewise constant control. Then there exists an interval $[0,T]$ and a piecewise constant control $U : [0,T] \to \Omega$ such that the solution to

$$\dot{X}(t) = U(t)X(t), \quad X(0) = I$$

on $[0,T]$ satisfies $X(T) = B$. But this is equivalent to saying that $\Phi_U(0,T) = B$. Thus we see that the first equality holds. Let $0 = t_0, t_1 < \ldots < t_n = T$ be a partition of $[0,T]$ such that U has constant value A_i on each (t_{i-1}, t_i) for $1 \leq i \leq n$. Since the solution of the preceding fundamental control equation for the constant control A_i is given by $\exp(tA_i)$, we have by Exercise 3.2.2(ii) and induction that

$$\Phi_U(t,0) = \exp(tA_i)\exp((t_{i-1} - t_{i-2})A_{i-1}) \cdots \exp(t_1A_1) \text{ for } t_{i-1} \leq t \leq t_i.$$

Since this is a finite product of members of $\exp(\mathbb{R}^+\Omega)$, we conclude that $S_{pc}(\Omega) \subseteq \langle \exp(\mathbb{R}^+\Omega)\rangle_{sgp}$.

Conversely given a member $\prod_{i=1}^n \exp(t_iA_i)$ where each $t_i \geq 0$ and each $A_i \in \Omega$, define a control $U : [0,T] \to \Omega$, where $T = \sum_{i=1}^n t_i$ by $U(t) = A_i$ for $\sum_{j=1}^{i-1} < t \leq \sum_{j=1}^{i} t_i$ and $U(0) = A_1$. Using the techniques of thepreceding paragraph, one sees that the solution of the fundamental control equation with initial condtion I has $\Phi(T,0)$ equal to the given product. ∎

We call $S_{pc}(\Omega)$ the *semigroup infinitesimally generated by* Ω.

Proposition 34. *For a control set Ω, $S(\Omega)$ is the closure of the reachable set from the identity.*

Proof. Let $B \in \mathcal{R}_I$, the reachable set from the identity. Then $B = X(T)$ for the solution of

$$\dot{X}(t) = U(t)X(t), \quad X(0) = I$$

for some bounded control $U : [0, T] \to \Omega$. Let Ω_1 be a bounded subset of Ω containing the image of U. By Proposition 31 $U_n \xrightarrow{w} U$ for some sequence of piecewise constant functions in $\mathcal{U}(\Omega_1, T)$. By Proposition 30, $\Phi_{U_n}(0, T) \to \Phi_U(0, T) = B$ and by the preceding lemma $\Phi_{U_n}(0, T) \in \langle \exp(\mathbb{R}^+\Omega)\rangle_{sgp}$ for each n. Hence $B \in \overline{\langle \exp(\mathbb{R}^+\Omega)\rangle}_{sgp} = S(\Omega)$. Therefore the reachable set from I and hence its closure is contained in $S(\Omega)$. Conversely, again using the preceding lemma,

$$S(\Omega) = \overline{\langle \exp(\mathbb{R}^+\Omega)\rangle}_{sgp} = \overline{S_{pc}(\Omega)} \subseteq \overline{\mathcal{R}_I}.$$

∎

Corollary 35. *Let Ω be a control set in $M_n(\mathbb{R})$ and S be a closed subsemigroup of $GL_n(\mathbb{R})$. Then $\Omega \subseteq \mathcal{L}(S)$ if and only if $\mathcal{R}_I \subseteq S(\Omega) \subseteq S$. In particular, $S(\Omega)$ is the smallest closed subsemigroup containing the reachable set from the identity, and $\mathcal{L}(S(\Omega))$ is the largest set $\widetilde{\Omega}$ satisfying $S(\widetilde{\Omega}) = S(\Omega)$.*

Proof. Suppose that $\Omega \subseteq \mathcal{L}(S)$. Then $\exp(\mathbb{R}^+\Omega) \subseteq S$. Since S is closed semigroup, $\overline{\langle \exp(\mathbb{R}^+\Omega)\rangle}_{sgp} \subseteq S$. The desired conclusion now follows from Lemma 33 and Proposition 34.

Conversely suppose that $\mathcal{R}_I \subseteq S$. Then $\exp(tA) \in \mathcal{R}_I \subseteq S$ for all $A \in \Omega$ and $t \geq 0.$, so $\Omega \subseteq \mathcal{L}(S)$.

By the preceding, $\Omega \subseteq \mathcal{L}(S(\Omega))$ (let $S = S(\Omega)$). Thus, again from the preceding, the reachable set from I is contained in $S(\Omega)$. Any other closed semigroup S containing $\mathcal{R}_I$ must contain $\exp(\mathbb{R}^+\Omega)$ and hence the closure of the semigroup it generates, which is $\overline{\langle \exp(\mathbb{R}^+\Omega)\rangle}_{sgp} = S(\Omega)$. On the other hand if $S(\widetilde{\Omega}) = S(\Omega)$, then we have just seen that $\widetilde{\Omega} \subseteq \mathcal{L}(S(\widetilde{\Omega})) = \mathcal{L}(S(\Omega))$.
∎

The next corollary involves only a mild modification of the preceding proof and is left as an exercise.

Corollary 36. *Let Ω be a control set in $M_n(\mathbb{R})$ and G a matrix group in $GL_n(\mathbb{R})$. The $\Omega \subseteq \mathfrak{g}$ if and only if $\mathcal{R}_I \subseteq S(\Omega) \subseteq G$. In particular, $G(\Omega)$ is the smallest matrix group containing the reachable set from the identity, and $\Omega \subseteq \mathcal{L}(G(\Omega)$.*

Exercise 3.3.1. Prove the preceding corollary.

Corollary 37. *If $S_{pc}(\Omega)$ is closed for a control set Ω, then $S_{pc}(\Omega) = S(\Omega) = \mathcal{R}_I$. In particular, every elment in the reachable set from I is reachable by a piecewise constant control.*

Proof. This follows from the last two lines of the proof of Proposition 34. ∎

3.4 The Sympletic Group: A Case Study

Recall that the sympletic group is given by

$$\mathrm{Sp}_{2n}(\mathbb{R}) = \left\{ \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathrm{GL}_{2n}(\mathbb{R}) : A^*C, B^*D \text{ are symmetric}, A^*D - C^*B = I \right\}$$

and its Lie algebra is given by

$$\mathfrak{sp}_{2n}(\mathbb{R}) = \left\{ \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in M_{2n}(\mathbb{R}) : D = -A^*,\ B, D \text{ are symmetric} \right\},$$

where in both cases $A, B, C, D \in M_n(\mathbb{R})$. Members of $\mathfrak{sp}_{2n}(\mathbb{R})$ are sometimes referred to as *Hamiltonian matrices*.

Recall that a symmetric matrix $A \in M_n(\mathbb{R})$ is *positive semidefinite*, written $A \geq 0$, if it satisfies $\langle x, Ax \rangle \geq 0$ for all $x \in \mathbb{R}^n$. A positive semidefinite matrix A is *positive definite*, written $A > 0$ if $\langle x, Ax \rangle > 0$ for all $x \neq 0$, or equivalently if it is invertible.

Exercise 3.4.1. (i) Characterize those diagonal matrices D satisfying $D \geq 0$ and $D > 0$.

(ii) For A symmetric and $P \in M_n(\mathbb{R})$, show that $A \geq 0$ implies $PAP^* \geq 0$. Show that the two are equivalent if P is invertible.

(iii) Use that fact that any symmetric matrix can be factorized in form $A = PDP^*$, where $P^* = P^{-1}$ is orthogonal and D is diagonal, to show that A is positive semidefinite (resp. definite) if and only if all its eigenvalues are non-negative (resp. positive).

(iv) Use the preceding factorization to show that any positive semidefinite matrix has a positive semidefinite square root.

(v) Show that the set of positive semidefinite matrices is closed in $M_n(\mathbb{R})$.

We consider the fundamental control system on $\mathrm{Sp}_{2n}(\mathbb{R})$ determined by

$$\Omega := \left\{ \left[\begin{array}{cc} A & B \\ C & D \end{array} \right] \in \mathfrak{sp}_{2n}(\mathbb{R}) : B, C \geq 0 \right\}.$$

We refer to members of Ω as non-negative Hamiltonian matrices. We also consider the set

$$\mathcal{S} = \left\{ \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathrm{Sp}_{2n}(\mathbb{R}) : D \text{ is invertible, } B^*D, CD^* \geq 0 \right\}.$$

Our goals in this section are to show that $\mathcal{S}$ is an infinitesimally generated semigroup with $\mathcal{L}(\mathcal{S}) = \Omega$. Then by Corollary 37, $\mathcal{S}$ will be the reachable set from the identity for the control system Ω.

Lemma 38. *If $P, Q \geq 0$ then $I + PQ$ is invertible. If $P > 0$ and $Q \geq 0$, then $P + Q > 0$.*

Proof. We first show that $I + PQ$ is injective. For if $(I + PQ)(x) = 0$, then

$$0 = \langle Q(x), (I + PQ)(x) \rangle = \langle Q(x), x \rangle + \langle Q(x), PQ(x) \rangle.$$

Since both latter terms are non-negative by hypothesis, we have that $0 = \langle Qx, x \rangle = \langle Q^{1/2}x, Q^{1/2}x \rangle$, and thus that $Q^{1/2}(x) = 0$. It follows that $0 = (I + PQ)(x) = x + PQ^{1/2}(Q^{1/2}x) = x$, and thus $I + PQ$ is injective, hence invertible.

The last assertion now follows easily by observing that $P + Q = P(I + P^{-1}Q)$. It follows that $P + Q$ is invertible and positive semidefinite, hence postive definite. ∎

We define

$$\begin{aligned} \Gamma^U &= \left\{ \begin{bmatrix} I & B \\ 0 & I \end{bmatrix} : B \geq 0 \right\}, \\ \Gamma^L &= \left\{ \begin{bmatrix} I & 0 \\ C & I \end{bmatrix} : C \geq 0 \right\}, \end{aligned}$$

We further define a group H of block diagonal matrices by

$$H = \left\{ \begin{bmatrix} A^* & 0 \\ 0 & A^{-1} \end{bmatrix} : A \in \mathrm{GL}_n(\mathbb{R}) \right\}.$$

The following lemma is straightforward.

Lemma 39. *The sets Γ^U and Γ^L are closed semigroups under composition. The semigroup Γ^U resp. Γ^L consists of all unipotent block upper (resp. lower) triangular operators contained in $\mathcal{S}$. The group H is closed in $\mathrm{GL}_{2n}(\mathbb{R})$ and consists of all block diagonal matrices in $Sp_{2n}(\mathbb{R})$. Furthermore, the semigroups Γ^U and Γ^L, are invariant under conjugation by members of H.*

Exercise 3.4.2. Prove Lemma 39.

Lemma 40. *We have that $\mathcal{S} = \Gamma^U H \Gamma^L$, Furthermore this "triple decomposition" is unique.*

Proof. Each member of $\mathcal{S}$ admits a triple decomposition of the form

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} (D^{-1})^* & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}.$$

The triple decomposition follows from direct multiplication (applying the equations $A^*D - C^*D = I$ and $B^*D = D^*B$ to see that the (1,1)-entry is A). Note further that if $B^*D = D^*B \geq 0$, then $BD^{-1} = (D^{-1})^*D^*BD^{-1} \geq 0$, and hence the first factor in the triple decomposition is in Γ^U, the block upper triangular matrices belonging to $\mathcal{S}$. Similar reasoning applies to showing the third factor belongs to Γ^L, the block lower triangular matrices belonging to $\mathcal{S}$, after noting $D^{-1}C = D^{-1}CD^*(D^{-1})^*$.

Conversely consider a product

$$\begin{bmatrix} D^{-1} + BD^* & BD^* \\ D^*C & D^* \end{bmatrix} = \begin{bmatrix} I & B \\ 0 & I \end{bmatrix} \begin{bmatrix} D^{-1} & 0 \\ 0 & D^* \end{bmatrix} \begin{bmatrix} I & 0 \\ C & I \end{bmatrix} \in \Gamma^U H \Gamma^L.$$

Then the (2,2)-entry in the product is precisely D^* and the middle block diagonal matrix in the factorization is determined. Multiplying the $(1,2)$-entry of the product on the right by $(D^*)^{-1}$ gives B and the $(2,1)$-entry on the left by $(D^*)^{-1}$ gives C. Hence the triple factorization is uniquely determined. Finally note that $(BD^*)^*D^* = DB^*D^*$ is positive semidefinite since B is (since the first block matrix is in Γ^U). Also $(D^*C)(D^*)^* = D^*CD$, which is positive semidefinite since C is. Thus the product block matrix satisfies the conditions to be in $\mathcal{S}$. ∎

We come now to an important theorem.

Theorem 41. *The set $\mathcal{S}$ is a subsemigroup of $Sp_n(\mathbb{R})$.*

Proof. Let $s_1 = u_1h_1l_1$ and $s_2 = u_2h_2l_2$ be the triple decompositions for $s_1, s_2 \in \mathcal{S}$. Suppose that $l_1u_2 = u_3h_3l_3 \in \Gamma^U H \Gamma^L$. That

$$s_1s_2 = u_1h_1l_1u_2h_2l_2 = u_1h_1u_3h_3l_3h_2l_2 = [u_1(h_1u_3h_1^{-1})](h_1h_3h_2)[(h_2^{-1}l_3h_2)l_2]$$

is in $\Gamma_0^U H \Gamma^L$ then follows from Lemma 39. We observe that indeed

$$l_1u_2 = \begin{bmatrix} I & 0 \\ C_1 & I \end{bmatrix} \begin{bmatrix} I & B_2 \\ 0 & I \end{bmatrix} = \begin{bmatrix} I & B_2 \\ C_1 & I + C_1B_2 \end{bmatrix},$$

and that the $(4,4)$-entry is invertible by Lemma 38. We further have that $B_2^*(I + C_1B_2) = B_2^* + B_2^*C_1B_2 \in \mathcal{P}_0$ is positive semidefinite and $C_1(I + C_1B_2)^* = C_1 + C_1B_2C_1^*$ is positive semidefinite since C_1 and B_2 are. Thus l_1u_2 has the desired triple decomposition $u_3h_3l_3$ and $\mathcal{S}$ is a semigroup by Lemma 40. ∎

The semigroup $\mathcal{S}$ of the preceding theorem is called the *sympletic semigroup.*

Corollary 42. *The sympletic semigroup can be alternatively characterized as*

$$\mathcal{S} = \left\{ \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in Sp_{2n}(\mathbb{R}) : A \text{ is invertible, } C^*A \in \mathcal{P},\ BA^* \in \mathcal{P} \right\}.$$

Proof. Let $\mathcal{S}'$ denote the set defined on the righthand side of the equation in the statement of this corollary. We observe that

$$\Delta \begin{bmatrix} A & B \\ C & D \end{bmatrix} \Delta = \begin{bmatrix} D & C \\ B & A \end{bmatrix} \text{ for } \Delta = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}.$$

The inner automorphism $M \mapsto \Delta M \Delta : \mathrm{GL}(V_E) \to \mathrm{GL}(V_E)$ carries $\mathrm{Sp}_{2n}(\mathbb{R})$ onto itself (check that it preserves the defining conditions at the beginning of this section), interchanges the semigroups Γ^U and Γ^L, carries the group H to itself, and interchanges the semigroup $\mathcal{S}$ and the set $\mathcal{S}'$. Thus $\mathcal{S}'$ is a semigroup and

$$\mathcal{S}' = \Gamma^L H \Gamma^U \subseteq \mathcal{S}\mathcal{S}\mathcal{S} = \mathcal{S}.$$

Dually $\mathcal{S} \subseteq \mathcal{S}'$. ∎

On the set of $n \times n$-symmetric matrices $\mathrm{Sym}_n(\mathbb{R})$, we define the *natural partial order* (also called the *Loewner order*), by $A \leq B$ if $0 \leq B - A$, that is, if $B - A$ is positive semidefinite.

Lemma 43. *Set* $S^+ = \{g \in Sp_{2n}(\mathbb{R}) : \forall s \in Sp_{2n}(\mathbb{R}),\ s_{21}^* s_{22} \leq (gs)_{21}^*(gs)_{22}\}$. (*Note that the matrices being compared are symmetric by the definition of* $Sp_{2n}(\mathbb{R})$.) *Then*

(i) S^+ *is a subsemigroup;*

(ii) $\mathcal{S} \subseteq S^+$*;*

(iii) S^+ *is closed.*

Proof. (i) This follows from the transitivity of the partial order.

(ii) To show the inclusion, we take any member of $\mathcal{S}$, write it in its triple composition, show that each factor is in S^+, and then use part (i). For example for $P \geq 0$,

$$\begin{bmatrix} I & 0 \\ P & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} * & B \\ * & PB + D \end{bmatrix}$$

and $(B^*(PB+D)) - B^*D = B^*PB \geq 0$, and thus the lower triangular matrix determined by P is in S^+. Similar arguments hold for members of H and Γ^U.

(iii) This follows from the continuity of the algebraic operations and the closeness in the symmetric matrices of the set of positive semidefinite matrices. ∎

Exercise 3.4.3. Work out the other two cases of part (ii) of the proof.

Theorem 44. *The sympletic semigroup* $\mathcal{S}$ *is closed.*

Proof. Let $s \in \overline{\mathcal{S}}$. Then $s = \lim_n s_n$, where each $s_n \in \mathcal{S}$. Then $(s_n)_{21}^*(s_n)_{11} \to s_{21}^* s_{11}$, so $s_{21}^* s_{11} \geq 0$ by the closeness of the set of positive semidefinite matrices and Corollary 42. Similarly $s_{12} s_{11}^* \geq 0$.

By parts (ii) and (iii) of Lemma 43, we have that $s \in \overline{\mathcal{S}} \subseteq S^+$. Thus for $P > 0$, we have

$$sr := \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & P \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & AP + B \\ C & CP + D \end{bmatrix},$$

where it must be the case that $0 < P = P^*I \leq (AP + B)^*(CP + D)$. It follows that $(AP + B)^*(CP + D)$ is positive definite, hence invertible, and thus $CP+D$ is invertible. Since $s_n r \to sr$ and each $s_n r \in \mathcal{S}$, we conclude that $(sr)_{21}^*(sr)_{22} \geq 0$ and similarly $(sr)_{12}(sr)_{22}^* \geq 0$. We conclude that $sr \in \mathcal{S}$, and hence by Corollary 42, that $A = (sr)_{11}$ is invertible. It now follows from Corollary 42 that $s \in \mathcal{S}$. ∎

We next establish that the tangent set of the symplectic semigroup $\mathcal{S}$ is the set Ω introduced at the beginning of this section.

Proposition 45. *The symplectic semigroup $\mathcal{S}$ has Lie wedge*

$$\mathcal{L}(\mathcal{S}) = \Omega = \left\{ \begin{bmatrix} A & B \\ C & -A^* \end{bmatrix} : B, C \geq 0 \right\}.$$

Proof. First note that any $X \in \Omega$ can be uniquely written as a sum

$$X = \begin{bmatrix} A & B \\ C & -A^* \end{bmatrix} = \begin{bmatrix} 0 & B \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} A & 0 \\ 0 & -A^* \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ C & 0 \end{bmatrix} = U + D + L$$

of a strictly upper block triangular, a block diagonal, and a strictly lower block triangular matrix. Since $\exp(tU) = \begin{bmatrix} I & tB \\ 0 & I \end{bmatrix} \in \Gamma^U \subseteq \mathcal{S}$ for all $t \geq 0$, we conclude that $U \in \mathcal{L}(\mathcal{S})$, and similarly $L \in \mathcal{L}(\mathcal{S})$. Clearly $\exp(tD) \in H \subseteq \mathcal{S}$ for all t, so $D \in \mathcal{L}(\mathcal{S})$ also. Since $\mathcal{L}(\mathcal{S})$ is a cone, hence closed under addition, we have that $X \in \mathcal{L}(\mathcal{S})$. Thus $\Omega \subseteq \mathcal{L}(\mathcal{S})$.

Conversely suppose that $\exp(tX) \in \mathcal{S}$ for all $t \geq 0$. Using the triple decompositions of Lemma 40, we can write

$$\exp(tX) = U(t)D(t)L(t) \text{ for each } t \geq 0.$$

Differentiating both sides with respect to t and evaluating at 0 yields

$$X = \dot{U}(0) + \dot{D}(0) + \dot{L}(0).$$

Then $X_{12} = \dot{U}(0)_{12} = \lim_{t \to 0^+} U(t)_{12}/t \geq 0$, since for the triple decomposition of any member of $\mathcal{S}$ the upper triangular factor $U(t)$ has positive definite $(1,2)$-entry (see Lemma 40 and proof). In a similar fashion one argues that $X_{21} \geq 0$. ∎

Corollary 46. *The closed semigroup $S(\Omega)$ infinitesimally generated by Ω is contained in the sympletic semigroup $\mathcal{S}$. In particular, the reachable set $\mathcal{R}_I(\Omega) \subseteq \mathcal{S}$.*

Proof. Since $\Omega = \mathcal{L}(\mathcal{S})$, $\exp(\mathbb{R}^+\Omega) \subseteq \mathcal{S}$. Since $\mathcal{S}$ is a closed semigroup, $S(\Omega) = \overline{\langle \exp(\mathbb{R}^+\Omega) \rangle}_{sgp} \subseteq \mathcal{S}$. By Proposition 34 the reachable set from the identity is contained in $S(\Omega)$. ∎

Exercise 3.4.4. Show that $\Gamma^U, \Gamma^L \subseteq \exp(\Omega) \subseteq \mathcal{R}_I(\Omega)$.

Chapter 4

Optimality and Riccati Equations

4.1 Linear Control Systems

Basic to the "state-space" approach to control theory is the theory of (time-varying) *linear systems* on $\mathbb{R}^n$:

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \quad x(t_0) = x_0,$$

where $x(t), x_0 \in \mathbb{R}^n$, $u(\cdot)$ belongs to the class of measurable locally bounded "control" or "steering" functions into $\mathbb{R}^m$, $A(t)$ is an $n \times n$-matrix and $B(t)$ is an $n \times m$ matrix. The system is *time-invariant* or *autonomous* if the matrices A and B are constant.

We may apply formula (3.5) of Proposition 3.6 to obtain the solution to the linear control system as

$$x(t) = \Phi(t, t_0)x_0 + \int_{t_0}^{t} \Phi(t, s)B(s)u(s)ds, \tag{4.1}$$

where $\Phi(t, t_0)$ is the fundamental solution for

$$\dot{X}(t) = A(t)X(t), \ X(t_0) = I.$$

Linear systems may be regarded as special cases of the fundamental control systems on groups that we considered in the previous chapter. Indeed

set $y(t) = \begin{bmatrix} x(t) \\ 1 \end{bmatrix}$, a column vector in $\mathbb{R}^{n+1}$, and consider the differential equation

$$\dot{y}(t) \;=\; U(t)y(t),\; y(t_0) = \begin{bmatrix} x_0 \\ 1 \end{bmatrix}, \quad \text{where } U(t) = \begin{bmatrix} A(t) & B(t)u(t) \\ 0 & 0 \end{bmatrix}.$$

One can see directly that $y(t)$ is a fundamental control equation of the type studied in the previous section, and that $y(t)$ satisfies this differential equation if and only if $\dot{x}(t) = A(t)x(t) + B(t)u(t)$ and $x(t_0) = x_0$.

Exercise 4.1.1. Verify the last assertion.

Since the fundamental control equation has global solutions, by the preceding considerations, the same holds for linear control equations.

Corollary 47. *Linear control systems have global solutions for the class of measurable locally bounded controls.*

4.2 The Linear Regulator

The basic optimization problem for linear control is the "linear regulator" or "linear-quadratic" problem with linear dynamics on $\mathbb{R}^n$ given by

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \quad x(t_0) = x_0, \tag{LIN}$$

where $A(t)$ is $n \times n$, $B(t)$ is $n \times m$ and $u(t) \in \mathbb{R}^m$. Associated with each control function $u : [t_0, t_1] \to \mathbb{R}^m$ is a quadratic "cost"

$$\int_{t_0}^{t_1} [x(s)'Q(s)x(s) + u(s)'R(s)u(s)]ds + x(t_1)'Sx(t_1),$$

where $R(s)$ is positive definite and $Q(s)$ is symmetric on $[t_0, t_1]$ S is symmetric, and $t_0 < t_1$ are fixed. Note that in this section we use a prime to denote the transpose, since we want to use stars to denote optimal choices. The integral is called the "running cost" and the last term is the "end cost."

An *optimal control* is a control function $u^* : [t_0, t_1] \to \mathbb{R}^m$ that minimizes the cost over all possible controls $u(\cdot)$. Its corresponding solution, denoted $x^*(\cdot)$, is the *optimal trajectory.*

A crucial ingredient in the theory and solution of the linear regulator problem is the associated quadratic *matrix Riccati Differential Equation* on $[t_0, t_1]$:

$$\dot{P}(t) = P(t)B(t)R^{-1}(t)B'(t)P(t) - P(t)A(t) - A'(t)P(t) - Q(t), \quad P(t_1) = S. \tag{RDE}$$

Theorem 48. *Suppose there exists a solution $P(t)$ on $[t_0, t_1]$ to (RDE). Then the solution of*

$$\dot{x}(t) = \big(A(t) - B(t)R(t)^{-1}B(t)'P(t)\big)x(t), \quad x(t_0) = x_0$$

yields the optimal trajectory $x^*(t)$ and the unique optimal control is the "feedback" control given by

$$u^*(t) = -R(t)^{-1}B(t)'P(t)x^*(t).$$

The minimal cost is given by $V(t_0, x_0) = x_0'P(t_0)x_0$.

Proof. Pick any initial state x_0 and any control $u : [t_0, t_1] \to \mathbb{R}^m$. Let $x(t)$ be the corresponding solution of (LIN). Taking the derivative with respect to t of $x(t)'P(t)x(t)$ (defined almost everywhere) , we obtain

$$\begin{aligned} \frac{d}{dt}x'Px &= (u'B' + x'A')Px + x'(PBR^{-1}B'P - PA - A'P - Q)x + x'P(Ax + Bu) \\ &= (u + R^{-1}B'Px)'R(u + R^{-1}B'Px) - u'Ru - x'Qx. \end{aligned}$$

Integrating we conclude

$$\begin{aligned} x(t_1)'P(t_1)x(t_1) - x_0P(t_0)x_0 &= \int_{t_0}^{t-1} (u + R^{-1}B'Px)'R(u + R^{-1}B'Px)(s)\,ds \\ &\quad - \int_{t_0}^{t_1} (u'Ru + x'Qx)(s)\,ds, \end{aligned}$$

and rearranging terms and noting $P(t_1) = Q$, we obtain

$$\begin{aligned} \text{cost} &= x(t_1)'Qx(t_1) + \int_{t_0}^{t_1} (u'Ru + x'Qx)(s)\,ds \\ &= x_0'P(t_0)x_0 + \int_{t_0}^{t-1} (u + R^{-1}B'Px)'R(u + R^{-1}B'Px)(s)\,ds. \end{aligned}$$

Since $R(s) > 0$ for all s, we conclude that the unique minimum cost control is the one making $u(s) + R^{-1}(s)B(s)'P(s)x(s) \equiv 0$, and in this case the minimal cost is $x_0'P(t_0)x_0$. ∎

4.3 Riccati Equations

We turn now to Ricatti equations.

Definition 49. A *matrix Riccati equation* is a differential equation on the vector space $\mathrm{Sym}_n(\mathbb{R})$ of symmetric $n \times n$-matrices of the form

$$\dot{K}(t) = R(t) + A(t)K(t) + K(t)A'(t) - K(t)S(t)K(t), \quad K(t_0) = K_0, \text{ (RDE)}$$

where $R(t), S(t), K_0$ are all in $\mathrm{Sym}_n(\mathbb{R})$.

There is a close connection between the fundamental group control equation on the sympletic group and the Riccati equation.

Lemma 50. *Suppose that $g(\cdot)$ is a solution of the following fundamental control equation on $Sp_{2n}(\mathbb{R})$ on an interval $\mathbb{I}$:*

$$\dot{g}(t) = \begin{bmatrix} A(t) & R(t) \\ S(t) & -A'(t) \end{bmatrix} \begin{bmatrix} g_{11}(t) & g_{12}(t) \\ g_{21}(t) & g_{22}(t) \end{bmatrix}, \quad R(t), S(t) \in \mathrm{Sym}_n(\mathbb{R}).$$

If g_{22} is invertible for all $t \in \mathbb{I}$, then $K(t) := g_{12}(t)(g_{22}(t))^{-1}$ satisfies

$$\dot{K}(t) = R(t) + A(t)K(t) + K(t)A'(t) - K(t)S(t)K(t),$$

on $\mathbb{I}$. Furthermore, if $g(t_0) = \begin{bmatrix} I & K_0 \\ 0 & I \end{bmatrix}$ for some $t_0 \in \mathbb{I}$, then $K(t_0) = K_0$.

Proof. Using the product rule and the power rule for inverses and the equality of the second columns in the fundamental control equation, we obtain

$$\begin{aligned} \dot{K} &= \dot{g}_{12}(g_{22})^{-1} - g_{12}g_{22}^{-1}\dot{g}_{22}g_{22}^{-1} \\ &= (Ag_{12} + Rg_{22})g_{22}^{-1} - K(Sg_{12} - A'g_{22})g_{22}^{-1} \\ &= AK + R - KSK + KA'. \end{aligned}$$

The last assertion is immediate. ∎

Corollary 51. *Local solutions exist for the Riccati equation (RDE) around any given intial condition.*

Proof. Global solutions exist for the fundamental control equation with initial condition $g(t_0) = \begin{bmatrix} I & K_0 \\ 0 & I \end{bmatrix}$ and the $g_{22}(t)$-entry will be invertible in some neighborhood of t_0 (by continuity of the determinant). Now apply the previous theorem. ∎

Our earlier results on the sympletic semigroup lead to a semigroup-theoretic proof of the following global existence result concerning the Riccati equation.

Theorem 52. *The Riccati equation*

$$\dot{K}(t) = R(t) + A(t)K(t) + K(t)A'(t) - K(t)S(t)K(t), \quad K(t_0) = K_0$$

has a solution consisting of positive semidefinite matrices for all $t \geq t_0$ if $R(t), S(t) \geq 0$ for all $t \geq t_0$ and $K_0 \geq 0$.

Proof. Let $g(t)$ be a global solutions for the fundamental control equation

$$\dot{g}(t) = \begin{bmatrix} A(t) & R(t) \\ S(t) & -A'(t) \end{bmatrix} \begin{bmatrix} g_{11}(t) & g_{12}(t) \\ g_{21}(t) & g_{22}(t) \end{bmatrix}, \; g(0) = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}.$$

Since by Corollary 46 the reachable set $\mathcal{R}_I(\Omega)$ is contained in the sympletic semigroup $\mathcal{S}$, and since for all t, $R(t), S(t) \geq 0$ implies the coefficient matrices in the fundamental control equation are all in Ω, we conclude that $g(t) \in \mathcal{S}$ for all $t \geq 0$. Set

$$h(t) := g(t - t_0) \begin{bmatrix} I & K_0 \\ 0 & I \end{bmatrix} :$$

since the fundamental control equation is right-invariant, $h(t)$ is again a solution, and $h(t) \in \mathcal{S}$ for all $t \geq 0$, since we have multiplied through by a member of the semigroup $\mathcal{S}$. Hence by definition of $\mathcal{S}$, $h_{22}(t)$ is invertible for all $t \geq t_0$. Thus by Lemma 50 $K(t) := h_{12}(t)(h_{22}(t))^{-1}$ defines a solution of the given Riccati equation for all $t \geq 0$, , and $K(t_0) = K_0 I^{-1} = K_0$.

Finally we verify that all $K(t)$ are positive semidefinite for $t \geq t_0$. Since $h(t) \in \mathcal{S}$, we have $(h_{12}(t))' h_{22}(t) \geq 0$. Thus

$$0 \leq (h_{22}^{-1}(t))'(h_{12}(t)' h_{22}(t)) h_{22}^{-1}(t) = (h_{12}(t) h_{22}^{-1}(t))' = h_{12}(t) h_{22}^{-1}(t) = K(t).$$

∎

The next corollary gives the more common setting in which solutions of the Riccati solution are considered.

Corollary 53. *The Riccati equation*

$$\dot{P}(t) = P(t)S(t)P(t) - A(t)P(t) - P(t)A'(t) - R(t), \quad P(t_1) = P_0$$

has a solution consisting of positive semidefinite matrices on the interval $[t_0, t_1]$ if $S(t), R(t) \geq 0$ for all $t_0 \leq t \leq t_1$ and $P_0 \geq 0$.

Proof. The equation

$$\dot{K}(t) = -K(t)S(t_1-t)K(t) - K(t)A(t_1-t) - A'(t_1-t)K(t) - R(t_1-t), \;\; P(0) = P_0$$

has a solution for all $0 \leq t \leq t_1 - t_0$ by the previous theorem, and $P(t) := K(t_1 - t)$ is then the desired solution of the differential equation given in this corollary. ∎

The next corollary is immediate from the results of this section and Theorem 48.

Corollary 54. *The Riccati equation associated with the linear regulator problem of Section 4.2 has an optimal solution as given in Theorem 48 if* $R(t) > 0$ *and* $Q(t) \geq 0$ *on* $[t_0, t_1]$ *and* $S \geq 0$.

Exercise 4.3.1. Calculate the feedback control that minimizes $\int_0^T (x_1^2 + u^2)dt$ for the system

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = u, \quad x(0) = x_0.$$

(Hint: Set up the Riccati equation that needs to be solved, consider the corresponding fundamental control equation $\dot{g} = \Gamma g$ on the group $\mathrm{Sp}_4(\mathbb{R})$, and find the fundamental solution $\exp(t\Gamma)$ by Fulmer's method. Set $K = g_{12}(g_{22})^{-1}$ and reparametrize to satisfy the appropriate initial conditions.)

4.4 Lifting Control Problems: An Example

In this section we consider an example of how control problems may sometimes be "lifted" to fundamental control problems on groups, which allows group machinery to be used in their study. The example also illustrates how certain problems in geometry or mechanics can be reinterpreted as control problems to allow application of the methods of control theory.

Let $\gamma(t)$ be a smooth curve in the plane $\mathbb{R}^2$ parameterized by arc length. The geometric study of the curve can be lifted to the group of rigid motions of $\mathbb{R}^2$ (the group generated by the orthogonal linear maps and the translations). We first assign a positively oriented frame v_1, v_2 to each point of the curve γ as follows. We set $v_1(t) = d\gamma/dt$, the tangent vector to the curve at $\gamma(t)$. Since γ is parameterized by arc length, we have that v_1 is a unit vector. Let v_2 denote the unit vector perpendicular to v_1 obtained by rotating v_1 counterclockwise by 90°. We say that v_2 is positively oriented with respect

to v_1 and call $(v_1(t), v_2(t))$ the *moving frame along* γ. Because $v_1(t)$ is a unit vector function, its derivative $\dot{v}_1(t)$ is perpendicular to $v_1(t)$, and hence a scalar multiple of $v_2(t)$. Hence there exists a scalar function $k(t)$, called the *curvature* of γ, such that

$$\dot{\gamma}(t) = v_1(t), \qquad \dot{v}_1(t) = k(t)v_2(t), \qquad \dot{v}_2(t) = -k(t)v_1(t)$$

One way of seeing the last equality is to rotate the curve γ by 90° counterclockwise and apply these same computations to the rotated curve. The preceding system of differential equations is called the *Serret-Frenet* differential system.

Exercise 4.4.1. (i) Show that if $v(t)$ is a differentiable unit vector function, then $v(t) \cdot \dot{v}(t) = 0$ for all t. (ii) Why does the minus sign appear in the last of the three displayed equations?

The moving frame along γ can be expressed by a rotation matrix $R(t)$ that rotates the frame $(v_1(t), v_2(t))$ to the standard basis (e_1, e_2), that is, $R(t)v_i = e_i$ for $i = 1, 2$. The matrix $R(t)$ has as its rows the representations of $v_1(t), v_2(t)$ in the standard basis. The curve $\gamma(t)$ along with its moving frame can now be represented as an element $g(t)$ in the motion group G_2 of the plane, a group of 3×3 matrices of the block form

$$g = \begin{bmatrix} R & 0 \\ \gamma & 1 \end{bmatrix},$$

with γ a row vector in $\mathbb{R}^2$ and $R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ a rotation matrix in $SO_2(\mathbb{R})$. The Serret-Frenet differential equations then convert to a fundamental equation on the group G_2:

$$\dot{g}(t) = \begin{bmatrix} 0 & k(t) & 0 \\ -k(t) & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} g(t).$$

Exercise 4.4.2. Verify that the rows of the preceding matrix differential equation are indeed equivalent to the three Serret-Frenet equations.

Exercise 4.4.3. Calculate the exponential solution of the preceding equation with initial condition I at time 0 if we assume the scalar function $k(t)$ is a constant function.

Exercise 4.4.4. Calculate the Lie algebra of the motion group G_2 on $\mathbb{R}^2$. (Hint: Show that each member of the group factors into a rotation times a translation, compute the Lie algebra of each subgroup, and sum.)

If the curvature function $k(t)$ is now regarded as a control function, then the preceding matrix differential equation converts to a fundamental control system on the group G_2, and many classical variational problems in geometry become problems in optimal control. For example, the problem of finding a curve $\gamma(t)$ that will satisfy the given boundary conditions $\gamma(0) = a$, $\dot{\gamma}(0) = \dot{a}$, $\gamma(T) = b$, $\dot{\gamma}(T) = \dot{b}$ (where $\dot{a}$ and $\dot{b}$ denote given tangent vectors at a and b resp.), and will minimize $\int_0^T k^2(t)dt$ goes back to Euler, and its solutions are known as *elastica*. Solutions may be interpreted as configurations that a thin, flexible plastic rods will take if the ends of the rods are to be placed at two specified points and directions at those points. More recent is the "problem of Dubins" of finding curves of minimal length that connect $(a, \dot{a})$ and $(b, \dot{b})$ and satisfy the contraint $|k(t)| \leq 1$. This may be viewed as a "parking problem," trying to maneuvern a vehicle from one point pointed in one direction to another point pointed in another direction with the curvature limit indicating how sharply you can turn the steering wheel.

Chapter 5

Geometric Control

5.1 Submanifolds of $\mathbb{R}^n$

In this chapter we introduce some basic ideas of that approach to control theory that is typically referred to as geometric control. The principal idea is to use the framework and tools of differential geometry to study problems in control theory and related underlying theory.

A *chart for* $\mathbb{R}^n$ is a diffeormorphism $\phi : U \to V$ between two nonempty open subsets of $\mathbb{R}^n$. For U a nonempty open subset of $\mathbb{R}^n$ and $S \subseteq U$, we say that S is a *k-slice* of U if there exists a chart $\phi : U \to \mathbb{R}^n$ such that

$$\phi(S) = \{(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n) \in \phi(U) : \, x_j = 0 \text{ for } k+1 \leq j \leq n\}.$$

A subset $S \subseteq \mathbb{R}^n$ is called an *embedded k-submanifold* of $\mathbb{R}^n$ if for each $p \in S$, there exists U open in $\mathbb{R}^n$ such that $S \cap U$ is a k-slice of U. Such charts are called *slice charts for* S. We shall also refer to such an embedded k-submanifold simply as a *submanifold* of $\mathbb{R}^n$, with the k being understood.

The next exercise shows that without loss of generality one can use a wider class of maps as slice charts, since tranlations and linear isomorphisms are diffeomorphisms on $\mathbb{R}^n$.

Exercise 5.1.1. Suppose that $\phi : U \to \mathbb{R}^n$ is a chart, $S \subseteq U$.

(i) If $\phi(S) = \{(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n) \in \phi(U) : \, x_j = a_j \text{ for } k+1 \leq j \leq n\}$ for given $a_{k+1}, \ldots, a_n$, show that ϕ followed by some translation on $\mathbb{R}^n$ is a slice chart.

(ii) Suppose $\phi : U \to \mathbb{R}^n$ is a slice chart for S and $x \in S \cap U$. Show there exists a slice chart $\psi : U \to \mathbb{R}^n$ such that $\psi(x) = 0$. (Hint: Follow ϕ by an appropriate translation.)

(iii) If $\phi(S) = \{(x_1, \ldots, x_n) \in \phi(U) : (x_1, \ldots, x_n) \in V\}$ for some subspace V of $\mathbb{R}^n$, show that ϕ followed by some linear isomophism (which may be chosen to be an orthogonal map) is a slice chart.

Proposition 55. *A matrix group G of $n \times n$ matrices is a submanifold of $M_n(\mathbb{R}) = \mathbb{R}^{n^2}$.*

Proof. We employ some basic facts about matrix groups from Chapter 1. First of all there exists an open ball B about 0 in $M_n(\mathbb{R})$ such that $\exp|_B$ is a diffeomorphism onto $V := \exp(B)$. Let $\phi : V \to M_n(\mathbb{R})$ be the inversemap, (what we previously called the log map). Let $\mathfrak{g}$ be the Lie algebra of G. Then by Theorem 20, by choosing B smaller if necessary, we may assume that $\exp|_{B\cap\mathfrak{g}}$ is a homeomorphism onto $\exp(B) \cap G = V \cap G$. It follows that $\phi(G \cap V) = \{A \in \phi(V) : A \in \mathfrak{g}\}$. By Exercise 5.1.1(ii), we may assume that ϕ is a slice chart. Thus we have created a slice chart at 1. To create a slice chart at any other $g \in G$, we take $\phi \circ \lambda_{g^{-1}}$, where $\lambda_{g^{-1}}$ is left translation under multiplication by g^{-1}. Since this is a linear isomorphism on $M_n(\mathbb{R})$, it is a diffeomorphism. ∎

Exercise 5.1.2. Use the preceding proposition and its proof to find a slice chart at $1 = (1, 0)$ of the unit circle subgroup of $C^* = \mathrm{GL}_1(\mathbb{C})$.

For an arbitrary nonempty subset $A \subseteq \mathbb{R}^n$, we say that $f : A \to \mathbb{R}^m$ is *smooth* if for each $x \in A$ there exists a set U open in $\mathbb{R}^n$ and containing x and a C^∞-function $g : U \to \mathbb{R}^m$ such that $g(y) = f(y)$ for all $y \in U \cap A$. In particular, we can speak of smooth functions $f : M \to N$ between submanifolds.

Exercise 5.1.3. Show that $f : M \to N$ between two submanifolds is smooth if and only if $\psi \circ f \circ \phi^{-1}$ is smooth on its domain of definition for all slice charts ϕ of M and ψ of N.

The *tangent bundle* of $\mathbb{R}^n$ is the pair $(T\mathbb{R}^n, \pi)$ where $T\mathbb{R}^n = \mathbb{R}^n \times \mathbb{R}^n$ and $\pi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is projection into the first coordinate. We endow the fibers $T_p\mathbb{R}^n := \pi^{-1}(p) = \{p\} \times \mathbb{R}^n$ with the structure of a vector space by taking the usual vector space operations in the second coordinate. Geometrically we visualize members of $T_p\mathbb{R}^n$ as directed line segments with initial point p.

We recall certain facts from calculus about curves and tangent vectors. Let $\alpha\colon (a,b) \to \mathbb{R}^n$ be a (parameterized) curve in $\mathbb{R}^n$ suchh that α is C^1. Then to each point $\alpha(t)$ on the curve, there exists a vector $\big(\alpha(t), \dot\alpha(t)\big)$ *tangent to the curve at* $\alpha(t)$ (also called the *velocity vector at* $\alpha(t)$), where $\dot\alpha(t)$ is defined by

$$\dot\alpha(t) = \lim_{h\to 0} \frac{\alpha(t+h) - \alpha(t)}{h}.$$

If α has components $\alpha_1, \ldots, \alpha_n$, then

$$\dot\alpha(t) = \big(\dot\alpha_1(t), \ldots, \dot\alpha_n(t)\big).$$

If M is a submanifold of $\mathbb{R}^n$, we define

$$TM = \{(p,v) \in T\mathbb{R}^n : p \in M,\, v \;=\; \dot\alpha(0) \text{ for some smooth curve } \alpha : (a,b) \to M \text{ with } \alpha(0) = p\}.$$

A member $(p,v) \in TM$ is called a *tangent vector at* p. The set of all tangent vectos at p is denoted T_pM. The restriction of $\pi : T\mathbb{R}^n \to \mathbb{R}^n$ to TM is denoted π_M and (TM, π_M) is called the *tangent bundle of* M.

Exercise 5.1.4. Consider the unit circle as a submanifold of $\mathbb{R}^2$. Find a description of its tangent bundle.

Lemma 56. *Let M be a k-dimensional submanifold of $\mathbb{R}^n$, $p \in M$. If $\phi : U \to \mathbb{R}^n$ is a slice chart and $p \in U$, then $T_pM = (d\phi^{-1})_{\phi(p)}(\mathbb{R}^k)$, where $\mathbb{R}^k$ is identified with the subspace of $\mathbb{R}^n$ with the coordinates after the first k all being 0.*

Proof. Let α be a smooth curve in M with $alpha(0) = p$. Set $\gamma := \phi \circ alpha$. Then the image of γ is contained in $\mathbb{R}^k$, so $\dot\gamma(0) \in \mathbb{R}^k$. It follows from the chain rule that $\dot\alpha(0) = (d\phi^{-1}_{\phi(p)}(\dot\gamma(0))$. Conversely for any $v \in \mathbb{R}^k$, the curve $\beta(\phi(p) + tv)$ satisfies $(d\phi^{-1})_{\phi(p)}(v) = \dot\beta(0)$. ∎

Exercise 5.1.5. (i) Show that the restriction of $\pi : T\mathbb{R}^n \to \mathbb{R}^n$ to M, namely $\pi_M : TM \to M$, is a smooth map..

(ii) If $\alpha : (a,b) \to M$ is a smooth curve show that $(\alpha(t), \dot\alpha(t)) \in T_{\alpha(t)}M := \pi_M^{-1}(\alpha(t))$ for $t \in (a,b)$.

(iii) Show that the tangent space T_pM is closed under scalar multiplication (the operation taking place in the second coordinate).

(iv) Show that T_pM is closed under addition. (Hint: Use the preceding lemma.)

Corollary 57. *Let G be a matrix group in $\mathrm{GL}_n(\mathbb{R})$. Then $T_eG = \mathfrak{g}$ and $T_gG = \mathfrak{g}g$.*

Proof. By Proposition 55 and its proof, a chart on small neighborhoods U of I is given by $\phi(g) = \log g$, and thus $\phi^{-1} = \exp$ on some open ball B around 0. Since $d\exp_0 = id_{M_n(\mathbb{R})}$ by Lemma 3, we conclude from Lemma 56 that $T_eG = \mathfrak{g}$. Since there are charts at $g \in G$ of the form $\log \circ \rho_{g^{-1}}$ with inverse $\rho_g \circ \exp$ (where ρ_g is right translation under multiplication by g), we conclude from Lemma 56 that

$$T_gG = d(\rho_g \circ \exp)_0(\mathfrak{g}) = d(\rho_g)_e \circ d\exp_0(\mathfrak{g}) = \rho_g(\mathfrak{g}) = \mathfrak{g}g,$$

where we use the fact that $d(\rho_g)_e = \rho_g$ since the map ρ_g is linear by the distributive law. ∎

5.2 Vector Fields and Flows

In this section we review certain basic facts about vector fields and flows on submanifolds.

Definition 58. A *C^r-real flow* for the additive group of real numbers on a submanifold M is a C^r-function $\Phi : \mathbb{R} \times M \to M$ satisfying

(i) the *identity property* $\Phi(0, x) = 0.x = x$,

(ii) and the *semigroup property*

$$\Phi\big(s, \Phi(t, x)\big) = \Phi(s + t, x), \text{ or } s.(t.x) = (s + t).x. \tag{SP}$$

A real flow is also called a *dynamical system*, or simply a *flow*.

For a C^1-dynamical system on M and $x \in M$, the set $\mathbb{R}.x = \Phi(\mathbb{R} \times \{x\}) = \phi_x(\mathbb{R})$ is called the *trajectory* or *orbit* through x. The sets $\mathbb{R}^+.x$ and $\mathbb{R}^-.x$ are called the *positive* and *negative semitrajectories* resp.

Suppose that x lies on the trajectory of y (and hence the trajectories of x and y agree). Then $x = \phi_y(t) := \Phi(t, y)$ for some t, the *motion* ϕ_y parametrizes the trajectory, and with respect to this parametrization, there is a tangent vector or velocity vector at x. There is also a tangent vector at x for the parametrization by the motion ϕ_x at $t = 0$. The next proposition asserts that the same vector is obtained in both cases.

Proposition 59. *For $y = \Phi(-t, x)$, we have that*

$$\dot{\phi}_x(0) = \dot{\phi}_y(t) = \frac{\partial \Phi}{\partial t}(t, y),$$

provided both exist. Hence for a C^1-flow, the tangent vector at x is the same for the parametrization of the trajectory of x by any motion ϕ_y, where y belongs to the trajectory of x.

Proof. We have $x = \phi^t \circ \phi^{-t}(x) = \phi^t(y)$, so

$$\phi_x(s) = \Phi(s, x) = \Phi\big(s, \Phi(t, y)\big) = \Phi(s + t, y) = \phi_y(s + t).$$

The assertion of the proposition now follows from an easy application of the chain rule at $s = 0$. ∎

Definition 60. A *vector field* $\mathbf{X}$ on a submanifold M of $\mathbb{R}^n$ is a function from M to TM which assigns to each $x \in M$ a tangent vector at x, i.e., a tangent vector in T_xM. Thus it must be the case that for each $x \in M$, there exists an unique $f(x) \in \mathbb{R}^n$ such that

$$\mathbf{X}(x) = \big(x, f(x)\big).$$

The function f is called the *principal part* of the vector field. Conversely any $f\colon M \to \mathbb{R}^n$, $\mathbf{X}(x) := \big(x, f(x)\big)$ such that $(x, f(x)) \in T_xM$ for each x defines a vector field with principal part f. Thus it is typical in this context to give a vector field simply by giving its principal part. The vector field is said to be C^r if it is a C^r function from M to TM; this occurs if and only if f is C^r into $\mathbb{R}^n$

Again we recall the standard geometric visualization of vector fields by sketching tangent vectors at a variety of typical points.

A C^r-flow for $r \geq 1$ gives rise in a natural way to a vector field. One considers for each $x \in M$ the trajectory through x parameterized by ϕ_x, and takes the velocity vector at time $t = 0$. The resulting vector field has principal part given by

$$f(x) := \frac{\partial \Phi}{\partial t}(0, x) = \frac{d\phi_x}{dt}|_{t=0} = \lim_{h \to 0} \frac{\phi_x(h) - x}{h}. \tag{DS1}$$

Definition 61. The vector field with principal part f given by equation (DS1) is called the *tangent vector field* or *velocity vector field* or *infinitesimal generator* of the flow Φ.

Remark 62. In light of Proposition 3, the vector at x can be chosen by choosing the tangent vector to x, where the trajectory of x is parametrized by any motion ϕ_y for y in the trajectory.

Proposition 63. *Let f be the principal part of the tangent vector field to a locally Lipschitz dynamical system Φ. For $x \in M$, x is a fixed point of Φ if and only if $f(x) = 0$.*

Proof. Exercise. ∎

Conversely one wishes to start with a vector field on a k-dimensional submanifold and obtain a dynamical system. For this direction one needs to utilize basic existence and uniqueness theorems from the theory of differential equations. Thus for a vector field $\mathbf{X} : M \to TM$, with principal part f, we consider the autonomous ordinary differential

$$\frac{dx}{dt} = \dot{x} = f(x), \tag{DS2}$$

called the *differential equation of the vector field*. Associating with a vector field the corresponding differential equation given by (DS2) establishes a straightforward one-to-one correspondence between vector fields on M and differential equations of the form (DS2) on M.

There are two ways in which we may use standard existence and uniqueness theorems to establish that unique solutions of (DS2) exist, at least locally. First of all, we may be able to extend (DS2) to a system on some open subset of $\mathbb{R}^n$ containing M and apply standard existence and uniqueness theorems to the extended system. Secondly we may "transfer" the system, at least locally, via a slice chart to a differential equation of $\mathbb{R}^k$, solve the equation there and then "pull back" the solution. This approach has the advantage that it guarantees that the system evolves on the submanifold, and if the vector field is locally lipschitz, so that solutions are unique, we conclude that the solutions from either approach must agree.

If $M = \mathbb{R}^n$ and we set f_i to be the composition of f with projection into the i-th coordinate, then equation (DS2) may be alternatively written as a system of equations

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, \ldots, x_n) \\ \dot{x}_2 &= f_2(x_1, \ldots, x_n) \\ &\vdots \\ \dot{x}_n &= f_n(x_1, \ldots, x_n). \end{aligned} \tag{DS3}$$

Suppose that for each $x \in M$, there exists a unique solution $x(t) = \Phi(t, x)$ of (DS2) which is defined on all of $\mathbb{R}$ and satisfies $x(0) = \Phi(0, x) = x$. Then it is well-known that the uniqueness of solutions implies

$$\Phi\big(s, \Phi(t, x)\big) = \Phi(s + t, x) \quad \text{for all } s, t \in \mathbb{R},$$

and that $\Phi\colon \mathbb{R} \times M \to M$ is continuous. Thus Φ defines a dynamical system on M, called the *real flow defined by* f (more precisely, the real flow generated by the vector field with principal part f), and from the way that it is defined, it is clear that the corresponding velocity vector field is the original one with principal part f. If $M = \mathbb{R}^n$, then a sufficient condition for such global solutions to exist is that f satisfies a *global Lipschitz condition*, i.e., that there exists $k > 0$ such that

$$\|f(x) - f(y)\| \leq k\|x - y\| \quad \text{for all } x, y \in \mathbb{R}^n.$$

If f is C^1, then it is locally Lipschitz, and this is enough to guarantee the existence and uniqueness of an appropriately defined *local flow*. In particular, if there is a global flow associated with the vector field, then it is unique. This yields the following:

Theorem 64. *Suppose that a real flow $\Phi\colon \mathbb{R} \times M \to M$ has a locally Lipschitz (in particular C^1) infinitesimal generator. Then the flow determines and is determined by the unique solution $t \mapsto \Phi(t, x)\colon \mathbb{R} \to M$ of (DS2) with initial value x at $t = 0$.*

In general, we assume that the local flows associated with the vector fields we encounter are actually global, and concern ourselves with problems of globality only on a case-by-case basis. Restricting to global flows is not a great loss in generality (see *Stability Theory of Dynamical Systems* by N. P. Bhatia and G. P. Szegö, Springer, 1970, p. 7).

Exercise 5.2.1. Find a formula for the real flow Φ defined by the following differential equations on the real line $\mathbb{R}$:

(i) $\dot{x} = 1$.

(ii) $\dot{x} = x$.

Exercise 5.2.2. (i) For a flow $\Phi : \mathbb{R} \times X \to X$, the function $\phi^t : X \to X$ defined by $\phi^t(x) = \Phi(t, x)$ is called a *transition function*. Show that the set of all transition functions ϕ^t that arise in (i) the preceding exercise is the group of translation functions on $\mathbb{R}$.
(ii) Show that the union of the two sets of transition functions ϕ^t arising in (i) and (ii) of the preceding exercise generates (under composition) the group of all proper affine transformations on $\mathbb{R}$:

$$\mathrm{Aff}_0(\mathbb{R}) := \{x \mapsto ax + b \colon \mathbb{R} \to \mathbb{R} \,:\, a > 0,\ b \in \mathbb{R}\}.$$

Exercise 5.2.3. (i) Solve the differential equation $\dot{x} = x^2$ on $\mathbb{R}$. Show that the flow defined by this differential equation is only a local flow, not a global one.
(ii) Find a flow on $\mathbb{R}^\infty := \mathbb{R} \cup \{\infty\}$ that has $\dot{x} = x^2$ as its infinitesimal generator.

Exercise 5.2.4. Define the notion of a constant vector field on $\mathbb{R}^n$, and find the real flow defined by such a vector field.

Exercise 5.2.5. Verify that the Φ defined after equation (DS3) does indeed satisfy the semigroup property under the assumption of uniqueness of solutions.

Exercise 5.2.6. Suppose that the principal part of a vector field on (an open subset of) E is given by $f(x) = -\nabla F(x)$, where F is C^r for $r \geq 1$ and ∇F is the gradient vector field. The corresponding real flow is called a *gradient flow* and F is called a *potential function for the flow*. Find the corresponding gradient flow $\Phi \colon \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2$ for

(i) $F(x, y) = x^2 + y^2$;

(ii) $F(x, y) = y^2 - x^2$.

What can be said about the trajectories of the flow and the level curves of F in each case?

Exercise 5.2.7. Consider the dynamical system on $\mathbb{R}^3$ given in cylindrical coordinates by $\Phi\big(t, (r, \theta, z)\big) = (r, \theta + \omega t, z)$.

(i) Give in words a description of this flow.

(ii) Find in Cartesian coordinates the tangent vector field of this flow.

(iii) Review from calculus the notion of the curl of a vector field and find the curl of this vector field. How does it relate to the original flow?

Exercise 5.2.8. Suppose that a vector field is the tangent vector field for a C^1-flow Φ on (an open subset of) E and that the principal part of the vector field is given by f. Show that x is a fixed point for the flow if and only if $f(x) = 0$.

5.3 Geometric Control

In geometric control one models the states of a system by the points of a smooth manifold and the controls by vector fields. The vector fields may be interpreted as force fields, for example an electrically charged surface or a gravitational field, or a velocity field such as represented along the surface of a moving body of water or along a moving air stream. To influence the system by controls then means that one varies the force field or velocity field of the manifold (for example, as the rudders of a wing change the configuration of air flow over the wing).

More precisely, the *dynamics* of a geometric control system are given by a function

$$F : M \times \Omega \to TM,$$

where M is a submanifold of $\mathbb{R}^n$ with tangent space TM, Ω has a least the structure of a metric space, and F is function satisfying $F(\cdot, u)$ is a smooth vector field for each $u \in \Omega$ and $F : M \times \Omega \to TM$ is continuous.

A *control function* is a locally bounded measurable (meaning the inverse of open sets are measurable) function $u(\cdot)$ from some subinterval of $\mathbb{R}$ to Ω. The control determines a differential equation on the manifold M:

$$\dot{x}(t) = F(x(t), u(t)).$$

A solution of the differential equation is an absolutely continuous function $x(\cdot)$ defined on the domain of the control function satisfying $\dot{x}(t) = F(x(t), u(t))$ almost everywhere.

Exercise 5.3.1. For a geometric control system and $x_0 \in M$, define the set controllable to x_0 and the set reachable from x_0.

We consider the fundamental control systems on matrix groups that we have been studying from the perspective of geometric control. Let G be a

matrix subgroup of $\mathrm{GL}_n(\mathbb{R})$. For each $A \in \mathfrak{g}$, we define a smooth vector field on G by $\mathbf{X}_A(g) = Ag$. By Corollary 57 we have that $Ag \in T_gG$, so that X_A is indeed a vector field. The map $f(B) = AB$ is a linear map on all of $M_n(\mathbb{R})$ that extends $\mathbf{X}_A$, so $\mathbf{X}_A$ is smooth. If we define a flow on G by $\Phi(t, g) = \exp(tA)g$, then Φ is a smooth flow on G with infinitesimal generator $\mathbf{X}_A$.

Exercise 5.3.2. Verify that Φ is indeed a flow and that $\dot{\phi}_g(0) = Ag$ for all $g \in G$.

We let $\mathfrak{g} \subseteq M_n(\mathbb{R})$ be the Lie algebra of G, and equip $\mathfrak{g}$ with the relative euclidean metric. We let Ω be a nonempty subset of $\mathfrak{g}$ and define the dynamics of a geometric control system by

$$F : G \times \Omega \to TG, \quad F(g, A) = Ag.$$

Note that $F(\cdot, A) = \mathbf{X}_A$, a smooth vector field, and that F is continuous, even smooth, since matrix multiplication is polynomial. Note also that for a control function $U : \mathbb{R} \to \Omega$ the corresponding control differential equation

$$\dot{g}(t) = F(g(t), U(t)) = U(t)g(t)$$

reduces to the time varying linear equation on G with coefficients from $\Omega \subseteq \mathfrak{g}$, what we called previously the fundamental control equation on G. Thus our earlier control system on a matrix group may be viewed as an important special case of a geometric control system on a smooth manifold.

One advantage of geometric control approach is that all Lie groups, not just matrix groups, admit the structure of a smooth manifold and hence the constructions of this chapter leadeing to geometric control systems can be carried out for them as well. Thus the control theory developed for matrix groups in these notes can be extended to general Lie groups via the machinery of smooth manifolds, tangent bundles, and vector fields. This approach also allows the introduction of the tools of differential geometry in the study of control systems.