SRC TR 87-194

Introduction to Manifolds, Lie Groups, and Estimation on the Circle

by

M.R. James

Introduction to Manifolds, Lie Groups, and Estimation on the Circle

M. R. James[1]

February, 1985
Revised October, 1987

[1]Mathematics Department and Systems Research Center, University of Maryland, College Park, MD, 20742, USA.

Abstract

These *lecture notes* introduce some basic concepts concerning smooth manifolds and Lie groups, since they are often the natural mathematical setting for studying nonlinear estimation and control problems. After reviewing some multivariable calculus, we discuss manifolds, derivative maps, vector fields and flows. Then Lie groups and Lie algebras are defined and discussed. Finally, we discuss the paper by Lo and Willsky which treats nonlinear estimation problems on the circle. A number of examples and illustrations are provided.

Key words: Differential geometry, Lie groups, nonlinear systems, nonlinear estimation.

INTRODUCTION TO MANIFOLDS, LIE GROUPS

AND ESTIMATION ON THE CIRCLE

1 Introduction

These notes are intended to provide an introduction to estimation problems which can be formulated on abelian Lie groups. The opportunity is taken to discuss some basic concepts about smooth manifolds and Lie groups, since they are often the natural mathematical setting for solving non-linear estimation and control problems.

The non-linear problem of optimal estimation on the circle S^1 can be solved using linear filtering theory, together with some non-linear transformations. This was done by Lo and Willsky [1], and the results can be extended to problems on abelian Lie groups. Examples include FM demodulation, frequency stability, and single degree of freedom gyroscope analysis.

2 Multivariable Calculus

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be *differentiable* at $p \in \mathbb{R}^n$ if there exists a linear map $f_{*p} : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\triangle p \to 0} \frac{|f(p + \triangle p) - f(p) - f_{*p} \triangle p|}{| \triangle p|} = 0$$

The matrix representation of f_{*p} is the Jacobian $[\frac{\partial f_i}{\partial p_j}(p)]$, $i = 1, \ldots, m;\ j = 1, \ldots, n$. Suppose $f : \mathbb{R}^n \to \mathbb{R}$, and v a vector in $\mathbb{R}^n$. The *directional derivative* of f in direction v at p is given by

$$v_p(f) = f_{*p}v = v_1 \frac{\partial f}{\partial p_1}(p) + \cdots + v_n \frac{\partial f}{\partial p_n}(p).$$

Chain rule. Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^m \to \mathbb{R}^q$ are differentiable at $p, f(p)$ respectively. Then $g \circ f : \mathbb{R}^n \to \mathbb{R}^q$ is differentiable at p and

$$(g \circ f)_{*p} = g_{*f(p)} \circ f_{*p}.$$

We say that f is *smooth* (or C^∞) if you can differentiate f as many times as you like.

Let $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$ open. Say $f : U \to V$ is a *diffeomorphism* if

i. f is smooth,

ii. f has a smooth inverse $f^{-1} : V \to U$.

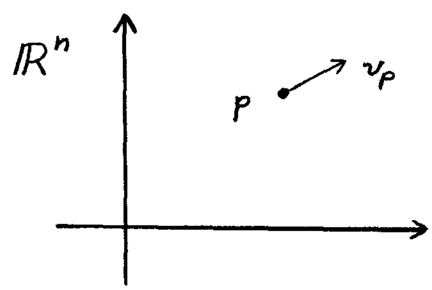
f is a *local diffeomorphism* at p if there exists a neighbourhood U of p such that

i. $f(U)$ is open,

ii. $f : U \to f(U)$ is a diffeomorphism.

Inverse Function Theorem Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is smooth. Then f is a local diffeomorphism at p if only if $f_{*p} : \mathbb{R}^n \to \mathbb{R}^m$ is an isomorphism. (So $n = m$.)

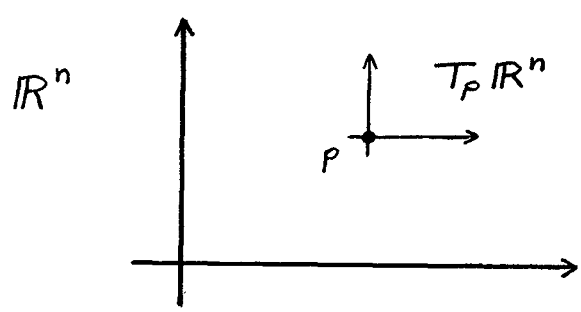
A *tangent vector* to $\mathbb{R}^n$ at p is a pair $v_p = (p, v) \in \mathbb{R}^n \times \mathbb{R}^n$, p is the point of attachment, and v is the vector part. Think of the arrow from p to $v + p$:



For all $p \in \mathbb{R}^n$, we define the *tangent space* to $\mathbb{R}^n$ at p to be the set

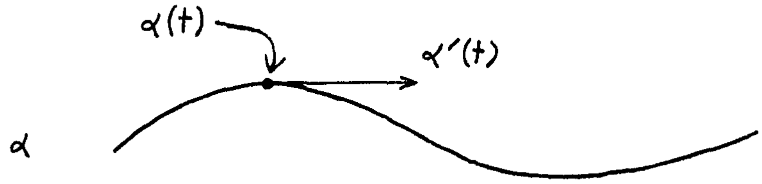
$$T_p(\mathbb{R}^n) = \{v_p : v \in \mathbb{R}^n\}.$$

This is an n-dimensional vector space.



A *smooth curve* is a smooth map $\alpha : I \to \mathbb{R}^n$, $t \mapsto \alpha(t)$ where $I \subset \mathbb{R}$ is an interval.

The *velocity vector* at t is the derivative $\alpha'(t) \in T_{\alpha(t)}(\mathbb{R}^n)$.



This is the tangent vector to α at t.

An *algebra* $\mathcal{A}$ is a vector space together with a multiplication $\bullet$ under which $\mathcal{A}$ is a ring. Scalar multiplication and $\bullet$ are related via:

$$\alpha(x \bullet y) = (\alpha x) \bullet y = x \bullet (\alpha y)$$

For any algebra $\mathcal{A}$, a linear map $T : \mathcal{A} \to \mathcal{A}$ satisfying

$$T(xy) = T(x)y + xT(y)$$

is called a *derivation*.

Here, let $\mathcal{A} = \{f : \mathbb{R}^n \to \mathbb{R} \text{ such that } f \text{ smooth}\}$. $\bullet$ is pointwise multiplication.

Every tangent vector $v_p \in T_p(\mathbb{R}^n)$ defines a differential operator $\mathcal{A} \to \mathbb{R}$ by

$$f \longmapsto v_p(f). \qquad \text{(directional derivative)}$$

This satisfies (product rule)

$$v_p(fg) = v_p(f)g(p) + f(p)v_p(g).$$

A *smooth vector field* is a map $V : p \longmapsto V_p \in T_p(\mathbb{R}^n)$.

The vector part depends smoothly on p.

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at p. We think of f_{*p} as a map between tangent spaces:

$$f_{*p} : T_p(\mathbb{R}^n) \longrightarrow T_{f(p)}(\mathbb{R}^m).$$

A vector field V defines a derivation $\mathcal{A} \to \mathcal{A}$

$$f \mapsto V[f]$$

where

$$V[f] : \mathbb{R}^n \to \mathbb{R}$$

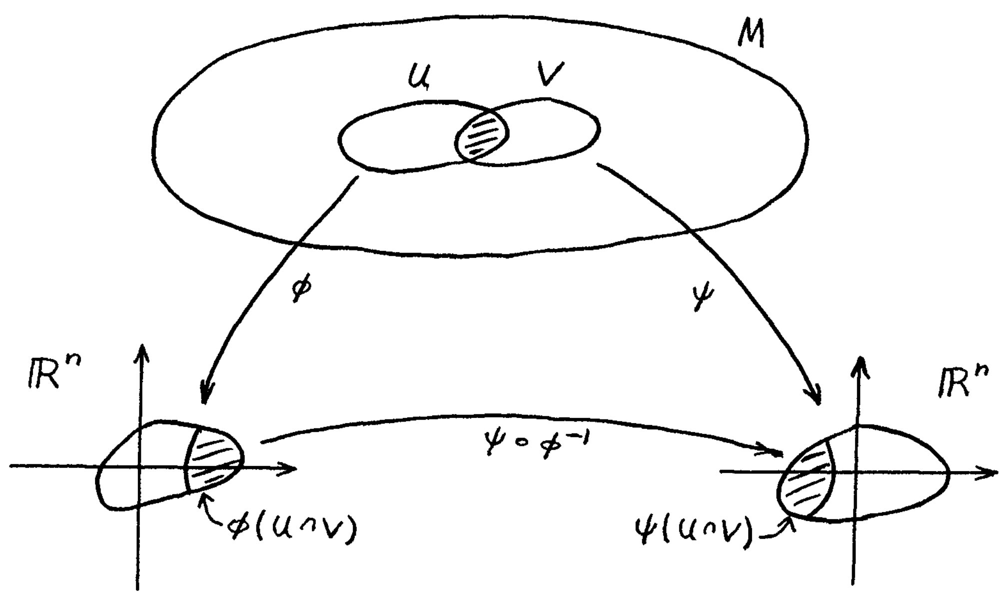
$$p \mapsto V_p(f) = f_{*p}(V_p)$$

So V acts on real valued functions, and when $V[f]$ is evaluated at a point p, we get the directional derivative of f in the direction V_p.

3 Smooth Manifolds

3.1 Definition

An n-dimensional *coordinate system* on a set M is a pair U, ϕ where $U \subset M$ and $\phi : U \to \mathbb{R}^n$ is injective and $\phi(U)$ is open in $\mathbb{R}^n$. (U, ϕ) is also called a *chart* on M.
Two coordinate systems (U, ϕ) and (V, ψ) are C^∞-*related* (or *compatible*) if the maps $\psi \circ \phi^{-1}$ and $\phi \circ \psi^{-1}$ are smooth (on $\phi(U \cap V)$ and $\psi(U \cap V)$ respectively). These maps "change coordinates".



An *atlas* on M is a collection of C^∞-related n-dimensional coordinate systems whose domains cover M. Two atlases are *compatible* if their union is an atlas. Compatability defines an equivalence relation on atlases.

An equivalence class of compatible atlases is called a *smooth structure* on M.

A *smooth n-dimensional manifold* is a set M with a smooth structure.

This defines a topology on M such that all coordinate systems for the smooth structure are homeomorphisms of open sets. Each point on the manifold has a neighbourhood homeomorphic to an open subset of $\mathbb{R}^n$. So manifolds look *locally* like $\mathbb{R}^n$. The global structure is in general very different.

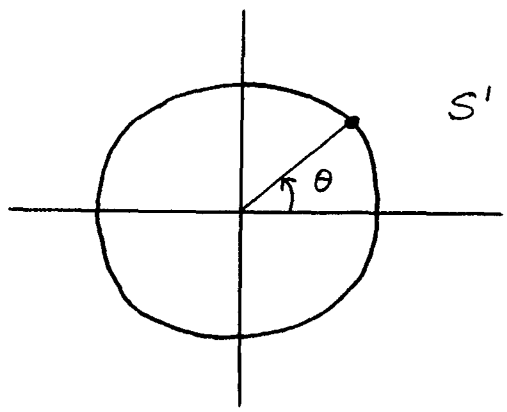
Examples

1. $\mathbb{R}^n$, n-dimensional *Euclidean space.*

2. S^n, n-dimensional *sphere.*

   $S^n = \{x \in \mathbb{R}^{n+1} : \|x\| = 1\}$

   Consider $S^1 = \{x \in \mathbb{R}^2 : \|x\| = 1\}$, the unit circle. (Also called the 1-dimensional torus T^1.)

   We use the sine and cosine functions to define a smooth structure for S^1:

   The charts $\left((-\frac{\pi}{2}, \frac{\pi}{2}),\ \sin\theta\right)$

   $\left((0, \pi),\ \cos\theta\right)$

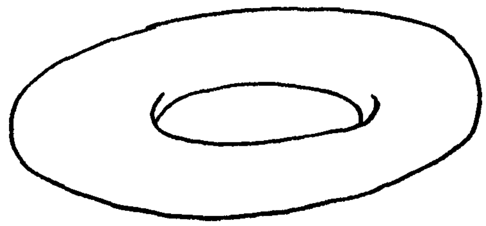
   $\left((\frac{\pi}{2}, \frac{3\pi}{2}),\ \sin\theta\right)$

   $\left((\pi, 0),\ \cos\theta\right)$

   are compatible and form an atlas.

3. T^n, the n- dimensional *torus*.

   $T^n = S^1 \times S^1 \times \cdots \times S^1 \quad (n - \text{times})$



T^2

4. $gl(n, \mathbb{R})$, real $n \times n$ matrices. This is an n^2-dimensional manifold homeomorphic to $\mathbb{R}^{n^2}$.

5. $Gl(n, \mathbb{R})$, *general linear group*, real $n \times n$ matrices with non-zero determinant. It is a manifold as an open subset of $\mathbb{R}^{n^2} \approx gl(n, \mathbb{R})$; the complement of the inverse image of 0 under det.

6. $SO(n, \mathbb{R})$, *special orthogonal group*, real $n \times n$ orthogonal matrices ($AA^T = I$) with det $= 1$.

Note There are topological manifolds that don't admit a smooth structure (a 10-dimensional one first found by Kervqire, 1960). Some topological manifolds admit more than one distinct smooth structure (e.g. S^7, Milnor).

Charts are often abbreviated $(x^1, \ldots, x^n)$. Suppose $p \in M$ and (U, ϕ) is a chart with $p \in U$. Then we could write $\phi = (x^1, \ldots, x^n)$ so that

$$\phi(p) = (x^1(p), \ldots, x^n(p)) \in \mathbb{R}^n.$$

The x^i are coordinate *functions*, and notation is often abused: $x^i = x^i(p)$ mean the coordinates of p.

3.2 Functions

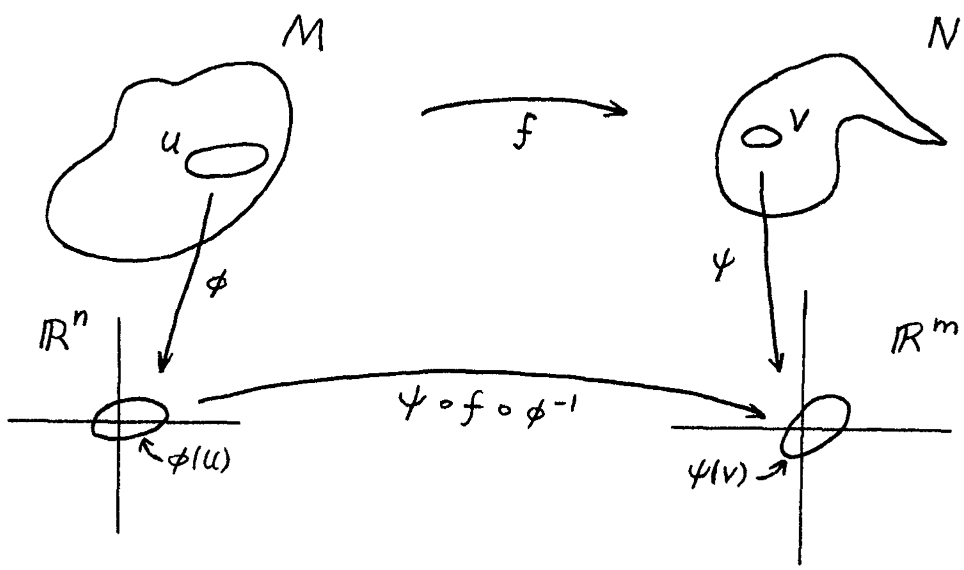
Consider the function

$$f : M \longrightarrow N$$

where M and N are smooth manifolds of dimension n and m respectively. Let (ϕ, U) and (ψ, V) be charts on M, N. The map

$$\psi \circ f \circ \phi^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

is called a *local representation* for f.



This expresses f in terms of local coordinates.

We say that f is *differentiable* at $p \in M$ if there exists a local representation which is differentiable at $\phi(p)$. This definition is intrinsic, that is, independent of choice of local representation.

A function $f : M \rightarrow N$ is *smooth* if it has a smooth representation at each point.

f is a *local diffeomorphism* at p if f has a local representation which is a local diffeomorphism at $\phi(p)$.

We say that f is a *diffeomorphism* if f is smooth and has a smooth inverse $f^{-1} : N \rightarrow M$. In such a case, M and N are said to be *diffeomorphic*.

3.3 Tangent Vectors and Tangent Spaces

We defined smooth manifolds without refering to any surrounding space. So we need intrinsic definitions for tangent vectors and tangent spaces. There are several ways for doing this. Perhaps the most obvious is to define a tangent vector in terms of charts and vectors in $\mathbb{R}^n$. All the definitions are related in a more or less obvious way. We give two definitions below.

Tangent Vectors as Equivalence Classes of Curves

Recall from Section 2 the notion of velocity vector. Many curves can have the same velocity vector at a point. Tangent vectors can be defined in terms of such curves. In Section 3.4 these ideas will be related to differential equations on manifolds.

Let $I \subset \mathbb{R}$ be an interval containing 0. A *smooth curve* is a smooth map $\alpha : \begin{array}{l} I \to M \\ t \mapsto \alpha(t). \end{array}$

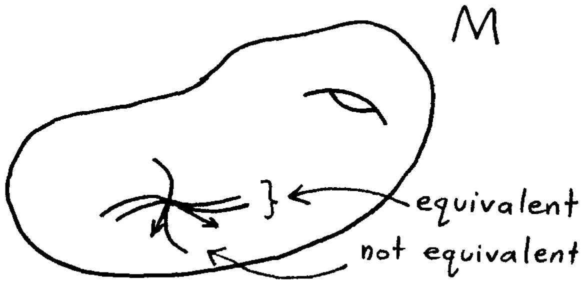
Define

$$C_p(M) = \{\alpha : I \to M : \alpha \text{ smooth and } \alpha(0) = p\}.$$

Curves in $C_p(M)$ are *equivalent* if their derivatives at p agree i.e. $\alpha, \beta \in C_p(M)$ are equivalent if there exist a chart (ϕ, U); $p \in U$, such that

$$(\phi \circ \alpha)'(0) = (\phi \circ \beta)'(0).$$

Define a *tangent vector* to be an equivalence class $[\alpha]_p$.

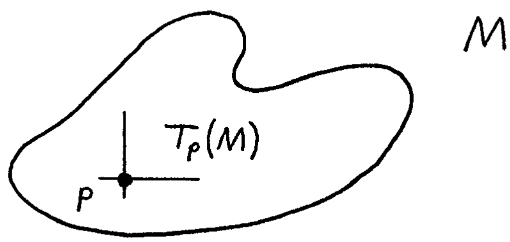


The *tangent space* at p to M, $T_p(M)$, is the set of all equivalence classes $C_p(M)/[]_p$.

The map $\tau_p : \begin{array}{l} T_p(M) \to \mathbb{R}^n \\ [\alpha]_p \mapsto (\phi \circ \alpha)'(0) \end{array}$

gives $T_p(M)$ a vector space structure, and is an isomorphism of vector spaces.

We can think of tangent spaces as n-dimensional vector spaces sitting at each $p \in M$.



Let $f : M \to N$ be a smooth map in a neighbourhood of p. The *derivative* map is defined to be

$$\begin{aligned} f_{*p} \;&:\; T_p(M) \to T_{f(p)}(N) \\ &[\alpha]_p \longmapsto [f \circ \alpha]_{f(p)} \end{aligned}$$

Chain rule. Suppose $f : M \to N$, $g : N \to L$ are differentiable at p, $f(p)$ respectively. Then $g \circ f : M \longrightarrow L$ is differentiable at p and

$$(g \circ f)_{*p} = g_{*f(p)} \circ f_{*p}.$$

Inverse Function Theorem Suppose $f : M \to N$ is smooth. Then f is a local diffeomorphism at p if and only if $f_{*p} : T_p(M) \to T_{f(p)}(N)$ is an isomorphism.

Remark. As you can see, calculus on manifolds is imported locally from $\mathbb{R}^n$. Much can be done in this more abstract setting.

Tangent Vectors as Differential Operators

Another very important way of viewing tangent vectors is as differential operators (recall section 1). This interpretation is used in Lie Algebra computations in non-linear filtering theory, see Marcus [2]. (Refer to Section 4).

Define $\mathcal{A}_p(M) = \{f : U \to \mathbb{R}; f \text{ is smooth for some open } U \subset M, p \in M\}$

Take $\alpha \in C_p(M)$. Define $\lambda_p^\alpha : \mathcal{A}_p(M) \to \mathbb{R}$ to be a directional derivative operator, by

$$\lambda_p^\alpha(f) = (f \circ \alpha)'(0) = [f \circ \alpha]_{f(p)} = f_{*p}([\alpha]_p).$$

Note that $\lambda_p^\alpha = \lambda_p^\beta$ if $\beta \in [\alpha]_p$. λ_p^α satisfies the product rule

$$\lambda_p^\alpha(fg) = \lambda_p^\alpha(f)g(p) + f(p)\lambda_p^\alpha(g).$$

We can define the *tangent space* at p to M, $\mathcal{T}_p(M)$, to be the set of all $\lambda : \mathcal{A}_p(M) \to I\!R$ satisfying the product rule. Elements of $\mathcal{T}_p(M)$ are called *tangent vectors.*

We have an isomorphism of vector spaces

$$\begin{aligned} T_p(M) &\longrightarrow \mathcal{T}_p(M) \\ [\alpha]_p &\longmapsto \lambda_p^\alpha \end{aligned}$$

(This isomorphism fails if f in the definition of $\mathcal{A}_p(M)$ is not smooth).

Also, we can readily construct an isomorphism $\mathcal{T}_p(M) \approx I\!R^n$. If $x^1, \ldots, x^n$ is a basis for $I\!R^n$, we can use this isomorphism to get a basis for $\mathcal{T}_p(M)$:

$$\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n} \;|_p$$

where $\frac{\partial}{\partial x^i} \;|_p$ is the directional derivative in the direction x^i evaluated at p. Hence if $v_p \in \mathcal{T}_p(M)$, and $f : M \to I\!R$ is differentiable at p, $v_p = \lambda_p^\alpha$, and

$$v_p(f) = \lambda_p^\alpha(f) = f_{*p}([\alpha]_p) = v_1 \frac{\partial f}{\partial x^1}(p) + \cdots + v_n \frac{\partial f}{\partial x^n}(p).$$

Tangent Bundle

All the tangent spaces can be put together to form another manifold, called the *tangent bundle* $T(M)$.

$$T(M) = \cup_{p \in M} T_p(M)$$

Projection map $\pi_M : \; T(M) \to M$

$$([\alpha]_p, p) \mapsto p$$

$$\pi_M^{-1}(p) = \text{ “fibre at } p\text{” } = T_p(M)$$

A 2n-dimensional smooth structure can be defined, under which $T(M)$ becomes a smooth manifold. For details, refer to one of the standard references.

The *derivative map* can now be viewed in a simple way. Let $f : M \to N$. Define the derivate map

$$\begin{aligned} f_* \;:\; & T(M) \to T(N) \\ & (p, [\alpha]_p) \mapsto (f(p), f_{*p}([\alpha]_p)) \end{aligned}$$

Chain rule $\qquad (g \circ f)_* = g_* \circ f_*$

3.4 Vector Fields and Flows

We are now in a position to define vector fields, in a similar manner as for $\mathbb{R}^n$, and to interpret trajectories (solution curves) of differential equations as curves whose tangents are defined by a vector field.

Vector Fields

A *smooth vector field* on a smooth manifold M is a smooth map

$$\begin{aligned} V \ : \ & M \to T(M) \\ & p \mapsto (p, V_p) \end{aligned}$$

assigning to each point a tangent vector at that point.

Define $\mathcal{A}(M) = \cup_{p\in M}\mathcal{A}_p(M)$. Then a smooth vector field V can be interpreted as a *derivation* on the algebra $\mathcal{A}(M)$, $V : f \mapsto V[f]$, where $V[f] : \ M \to \mathbb{R}$

$$p \mapsto V_p(f) = f_{*p}(V_p).$$

This again gives directional derivatives. In terms of local coordinates, let (ϕ, U) be a chart on M, $p \in M$. Then

$$V\lceil_U = f_1\frac{\partial}{\partial x^1} + \cdots + f_n\frac{\partial}{\partial x^n}$$

where $f_i : U \to \mathbb{R}$ are smooth and $\frac{\partial}{\partial x^i}$ are vector fields on U, giving directional derivatives in the direction x^i.

Flows

Suppose V is a smooth vector field on M, and I is an interval containing 0. We say that a curve $\alpha : I \to M$ is an *integral curve* (or *trajectory*) for V if α is smooth and

$$\alpha'(t) = V(\alpha(t)), \ \text{ for all } \ t \in I.$$

So the tangent to α at t is given by the vector field, in the same way that a differential equation determines the rates of change of its solutions.

Briefly, a *flow* on M is a function $F(t,p)$ defined on (possibly only a subset of) $\mathbb{R} \times M$ with values in M, such that for each p,

$$\alpha_p : t \mapsto F(t,p)$$

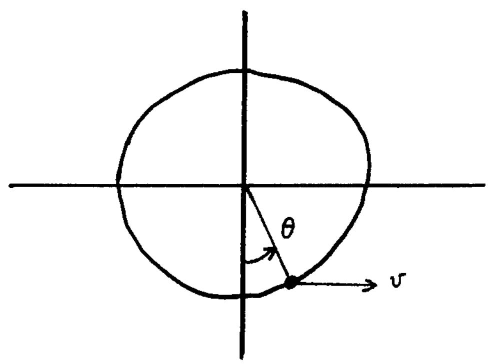
is a trajectory for V with initial value $\alpha_p(0) = p$. Think of a flow as a family of trajectories, indexed by points in M.

A smooth vector field is *complete* if the domain of F is all of $\mathbb{R} \times M$. Then for each t,

$$F_t : p \mapsto F(t, p)$$

is a diffeomorphism $M \to M$. The set $\{F_t : t \in \mathbb{R}\}$ is a group (under composition, $F_t \circ F_s = F_{t+s}$), called a 1-parameter group of diffeomorphisms on M.

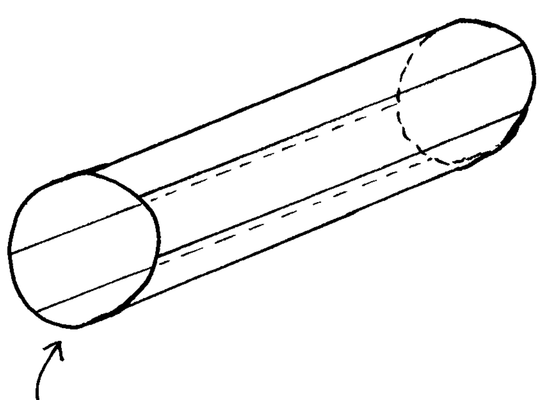
Example (Simple Pendulum)



The state of the pendulum is given by position $\theta \in S^1$

velocity $v \in \mathbb{R}$

The *phase space* (state space) is $S^1 \times \mathbb{R}$, a cylinder which is a 2-dimensional surface:



$$S^1 \times \mathbb{R} = \{(\theta, v) : -\pi < \theta \leq \pi, -\infty < v < \infty\}$$

The differential equation describing the motion of the pendulum is

$$\dot{\theta}(t) = v(t)$$

$$\dot{v}(t) \;=\; -\sin\theta(t)$$

The trajectories α are curves on the cylinder, and satisfy

$$\begin{aligned} \alpha'(t) = (\dot{\theta}(t), \dot{v}(t)) &= V(\theta(t), v(t)) \\ & (v(t), -\sin\theta(t)) \end{aligned}$$

where $V(\theta, v) = (v, -\sin\theta)$ is the vector field for the pendulum.

4 Lie Groups and Lie Algebras

In this section, Lie groups and Lie algebras are defined, and some basic relations discussed. A Lie group is a smooth manifold which is also a group, and has associated with it an algebra, called the Lie algebra, and there is a mapping between them, called the exponential map.

Definitions

A *Lie group* G is a smooth manifold with a compatible group structure, that is, the maps

$$\begin{aligned} &G \times G \to G \\ &(\sigma, \tau) \mapsto \sigma\,\tau \quad \text{(multiplication)} \end{aligned}$$

$$\begin{aligned} &G \to G \\ &\sigma \mapsto \sigma^{-1} \qquad \text{(inversion)} \end{aligned}$$

are smooth. (It is enough to show that the map $(\sigma, \tau) \mapsto \sigma\,\tau^{-1}$ is smooth). The identity element of G is denoted by e.

Examples (see end of this section): $\mathbb{R}^n, S^1, T^n, Gl(n, \mathbb{R}),\ SO(n, \mathbb{R})$.

A *Lie algebra* g over $\mathbb{R}$ is a real vector space of together with a bilinear operator (*bracket, commutator*)

$$[,] : g \times g \longrightarrow g$$

such that $\forall\, x, y, z \in g$

$$[x, y] = -[y, x] \quad \text{(anti-commutativity)}$$

$$[[x,y],z]+[[y,z],x]+[[z,x],y]=0 \quad \text{(Jacobi identity)}$$

We say that g is *abelian* if $[x,y] = 0\ \forall\ x,y \in g$. So in general Lie algebras are non-associative, and the bracket $[,]$ corresponds to the multiplication $\bullet$ discussed in Section 1.

Examples $\mathbb{R}^n, gl(n,\mathbb{R})$, vector fields.

Lie Brackets

The Lie bracket of two *matrices* $A,B \in gl(n,\mathbb{R})$ is the usual commutator

$$[A,B]=AB-BA,$$

making $gl(n,\mathbb{R})$ a Lie algebra.

Let $\mathcal{A}$ be an *algebra*, and $\mathcal{D}$ the algebra of derivations on $\mathcal{A}$. Given two derivations $S,T \in \mathcal{D}$, define their bracket

$$[S,T]=S\circ T-T\circ S,$$

this is also a derivation. So $\mathcal{D}$ is a Lie algebra, called the *derivation algebra* of $\mathcal{A}$.

Let $\mathcal{A}=\mathcal{A}(M), \mathcal{D}$ be the algebra of all *smooth vector fields* on M, viewed as derivations. If V,W are smooth vector fields, so is their bracket $[V,W]$; meaning

$$[V,W]_p(f)=V_p(Wf)-W_p(Vf).$$

(Recall that $Vf \equiv V[f]$ is a smooth function $M \to \mathbb{R}$; and $W_p(Vf)=(Vf)_{*p}(W_p)$, the directional derivative of Vf in the direction W_p.)

In terms of *local coordinates*, let V,W have representations with respect to a chart (ϕ,U)

$$V|_U=\sum_i v_i\frac{\partial}{\partial x^i} \text{ and } W|_U\sum_i w_i\frac{\partial}{\partial x^i}.$$

Then the bracket $[V,W]$ has a representation

$$[V,W]|_U=\sum_i\left(\sum_j\left(v_j\frac{\partial w_i}{\partial x^j}-v_j\frac{\partial v_i}{\partial x^j}\right)\right)\frac{\partial}{\partial x^i}.$$

Application [2] (Non-linear control systems)

Consider a *control system*, expressed in local coordinates

$$\dot{x}_t=f(x_t)+\sum_{i=1}^m u_t^i g_i(x_t)$$

(The RHS is the vector field) where $x_t \in M$, a smooth manifold; f, g_i are smooth vector fields on M; and u^i are control functions.

The Lie algebra $\mathcal{L} = \langle f, g_1, \ldots, g_m \rangle$ generated by the vector fields is called the *controllability Lie algebra* of the control system.

If $f(x) = Ax$, $g_i(x) = B_i x$, where A and B_i are matrices, the control system is *bilinear*

$$\dot{x}_t = Ax_t + \sum_{i=1}^{m} u_t^i b_i x_t.$$

Suppose V, W are *linear vector fields*, locally,

$$V_x = Ax\ ,\ \ W_x = Bx$$

where A, B are matrices. Then $[V, W] = [B, A]$, since

$$[V, W]_x = [As, Bx] = [BA - AB]x.$$

Left Invariant Vector Fields

Lie groups, Lie algebras and (left invariant) vector fields are now related.

Take $\sigma \in G$. Define *left translation* by σ to be the diffeomorphism

$$\begin{aligned} l_\sigma : \quad & G \to G \\ & \tau \mapsto \sigma\tau. \end{aligned}$$

A vector field V is said to be *left invariant* if for each $\sigma \in G, V$ is l_σ-related to itself, i.e.

$$(l_\sigma)_* \circ V = V \circ l_\sigma, \ \ \forall \sigma \in G.$$

This means that

$$(l_\sigma)_{*\tau}(V_\tau) = V_{\sigma\tau}, \ \ \forall \sigma, \tau \in G.$$

So left invariance is equivalent to the property that

$$(l_\sigma)_{*e}(V_e) = V_\sigma, \quad \forall \sigma \in G.$$

Thus a left invariant vector field is determined by its value at the identity e, V_e.

Denote by g the set of all left invariant vector fields on G. We have the following results ([3], page 25):

(i) g is a vector space, and the map

$$g \to T_e(G)$$
$$V \mapsto V_e$$

is an isomorphism. So $\dim g = \dim T_e(G) = \dim G$.

(ii) Left invariant vector fields are smooth.

(iii) The Lie bracket of two left invariant vector fields is itself a left invariant vector field.

(iv) g forms a Lie algebra under the Lie bracket operation on vector fields.

Thus we can associate a Lie algebra to each Lie group - namely the tangent space at the identity e, with the induced bracket. The converse is a difficult result; if g is a Lie algebra, then there is a simply connected Lie group which has g as its Lie algebra. (See [3], p.101).

Homomorphisms

If G, H are Lie groups, a map $\phi : G \to H$ is said to be a (*Lie group*) *homomorphism* if ϕ is smooth and a group homomorphism of the abstract groups. If ϕ is also a diffeomorphism, say ϕ is an *isomorphism.*

If g, h are Lie algebras, a map $\psi : g \to h$ is said to be a (*Lie algebra*) *homomorphism* if ψ is linear and

$$\psi[V,W] = [\psi(V), \psi(W)] \text{ for all } V,W \in g.$$

If in addition ψ is bijective, say ψ is an *isomorphism.*

Let G, H be Lie groups with Lie algebras g, h, and $\phi : G \to H$ a Lie group homomorphism. Then we can use the derivative map $\phi_{*e} : T_e(G) \to T_e(H)$ to define a Lie algebra homomorphism $d\phi : g \to h$, with $d\phi(V)(e) = \phi_{*e}(V_e)$.

Note A connected Lie group is abelian if only if its Lie algebra is abelian.

Exponential Map

Take a left-invariant vector field $V \in g$, and consider the integral curve α_v

$$\alpha_v'(t) = V(\alpha_v(t)); \quad \alpha_v(0) = e.$$

Left invariance implies $\alpha_v(t+s) = \alpha_v(t)\alpha_v(s)$, so that $\alpha_v : I\!R \to G$ is a Lie group homomorphism.

We can now define the *exponential map*

$$\begin{aligned} \exp \ : \ & g \to G \\ & V \mapsto \alpha_v(1). \end{aligned}$$

Properties of the exponential map include the following ([3], page 103).

(i) $\exp(tV) = \alpha_v(t), \qquad \forall t \in I\!R.$

(ii) $\exp(t_1 + t_2)V = \exp(t_1 V)\exp(t_2 V), \qquad \forall t_1, t_2 \in I\!R..$

(iii) $\exp(-tV) = (\exp(tV))^{-1}, \qquad \forall t \in I\!R.$

(iv) $\exp : g \to G$ is smooth, and $\exp_* : T_o(g) \to T_e(G)$ is the identity map. Hence by the inverse function theorem, exp gives a diffeomorphism of a neighbourhood of 0 in g onto a neighbourhood e of G.

(v) If $[V, W] = 0$ then

$$\exp(V + W) = \exp(V)\exp(W).$$

(vi) If $\phi : H \to G$ is a homomorphism, then the diagram commutes:

$$\begin{tikzcd} H \arrow[r, "\phi"] & G \\ h \arrow[u, "\exp"] \arrow[r, "d\phi"'] & g \arrow[u, "\exp"'] \end{tikzcd}$$

<u>Examples</u>

1. $I\!R^n$ is an abelian Lie group, under vector addition, with Lie algebra just $I\!R^n$.

2. S^1 is an abelian Lie group. The group structure is defined by considering S^1 as the quotient group

$$I\!R/2\pi Z$$

under addition. An element $\theta \in S^1$ is an equivalence class of real numbers, modulo 2π. So we can represent $\theta \in S^1$ as a number, or an angle,

$$\theta \in [-\pi, \pi)$$

This group structure is compatible with the smooth manifold structure, as follows. We want to show that for $\sigma, \tau \in S^1$ the map $(\sigma, \tau) \mapsto \sigma\tau^{-1}$ is smooth. Using our definition, $\sigma\tau^{-1}$ is given by

$$\sigma - \tau, \ \text{mod } 2\pi.$$

In terms of coordinates, this is

$$(\sigma, \tau) \mapsto \cos(\sigma - \tau) = \cos\sigma\cos\tau + \sin\sigma\sin\tau$$

or

$$(\sigma, \tau) \mapsto \sin(\sigma - \tau) = \sin\sigma\cos\tau - \sin\tau\cos\sigma,$$

and is clearly smooth.

Note that the circle S^1 can be represented as

$$\{z \in C : |z| = 1\} = \{e^{i\theta} : \theta \in [-\pi, \pi)\}.$$

Another representation is $S^1 = SO(2, I\!R) = SO(2)$, as can be seen by the correspondence (c.f. rotation)

$$\begin{array}{ccc} [-\pi, \pi) & \longleftrightarrow & SO(2) \\ \theta & \longleftrightarrow & \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \end{array}$$

The Lie algebra of $S^1 = SO(2)$ is $o(2, I\!R) \approx I\!R$ (see example 6 below). A basis is

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

the infinitesimal rotation matrix.

The exponential map is given by

$$\begin{array}{ccc} \exp : I\!R & \longrightarrow & S^1 \\ o(2) & \longrightarrow & SO(2) \\ \theta & \longmapsto & \exp(R\theta) \\ \begin{bmatrix} 0 & \theta \\ -\theta & 0 \end{bmatrix} & \longmapsto & \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \end{array}$$

Note S^1 and S^3 are the only spheres admitting a Lie group structure.

3. T^n is an abelian Lie group, with group structure induced from S^1. Lie algebra is $I\!R^n$.

4. Every connected abelian Lie group is of the form

$$I\!R^n \times T^m,$$

with Lie algebra $I\!R^{n+m}$.

5. $gl(n, I\!R)$ is a Lie algebra, with bracket given by the matrix commutator.

   $GL(n, I\!R)$ is a Lie group, under matrix multiplication.

   $gl(n, I\!R)$ is the Lie algebra of $GL(n, I\!R)$.

   The exponential map is given by exponentiation of matrices.

6. $SO(n, I\!R)$ is a Lie group, under matrix multiplication. Its Lie algebra is $o(n, I\!R)$; the real skew-symmetric matrices, with the matrix commutator. exp is the matrix exponential.

5 Estimation on The Circle

In this section, we discuss the paper by Lo and Willsky [1]. In the introduction, they give the following example. We wish to recover a process x from a transmitted signal

$$s(t) = e^{i\int_o^t x(s)ds}.$$

During transmission, the signal is corrupted by noise, so that the received signal is

$$z(t) = e^{i\,v(t)}\,s(t),$$

where v is a Brownian motion. Note that the processes s and z evolve on the circle $S^1 = \{|z| = 1\}$. In this example, the idea is to design an optimal filter to recover x from z. Another example to keep in mind is the problem of estimating angular orientation, from noisy measurements taken by a sensor in a gyroscope.

<u>Random Processes on $I\!R$ and S^1</u>

Let us work on a probability space $(\Omega, \mathcal{A}, P)$, and use any of the representations for S^1 discussed in Section 4.

Let C_1^s be the space of real valued continuous functions a defined on $[0, s]$ such that $a(0) = 0$, and $\mathcal{B}_1^s$ be the Borel σ-algebra of C_1^s.

Let C_2^s be the space of SO (2) valued continuous functions A defined on $[0, s]$ with $A(0) = I$, and $\mathcal{B}_2^s$ the Borel σ-algebra of C_2^s. (The topologies of C_1^s and C_2^s are defined by sup norms.)

Define a bijective $\mathcal{B}_1^s$-measurable function

$$\begin{aligned} J \ :\ & C_1^s \to C_2^s \\ & a \mapsto A \end{aligned}$$

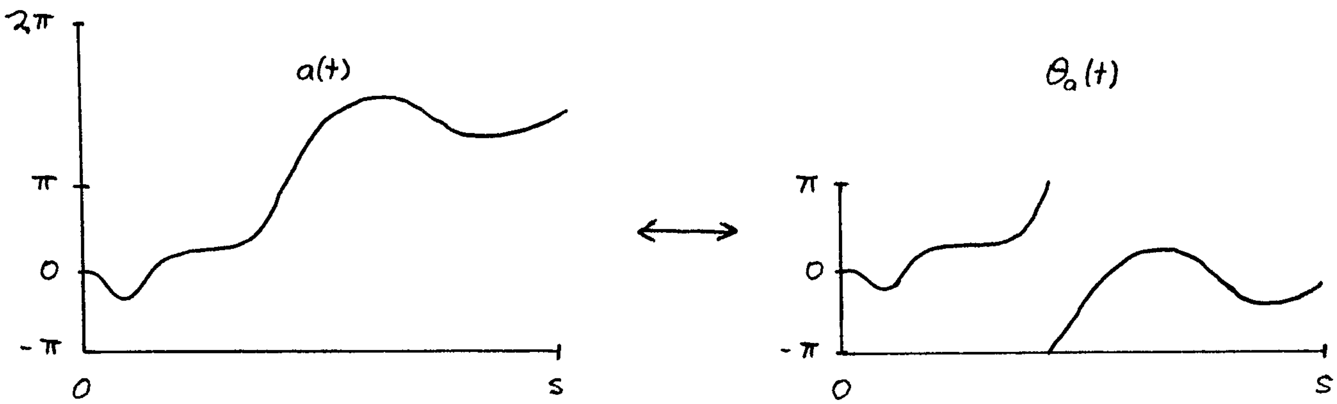
by

$$A(t) = (J_a)(t) = \exp(a(t)R) = \begin{bmatrix} \cos a(t) & \sin a(t) \\ -\sin a(t) & \cos a(t) \end{bmatrix}$$

for $t \in [0, s]$. Think of this as the map

$$a(t) \mapsto a(t) \mod 2\pi = \theta_a(t) \in [-\pi, \pi).$$

The bijectivity means that given a function a, we can construct a unique function A, and vice versa.



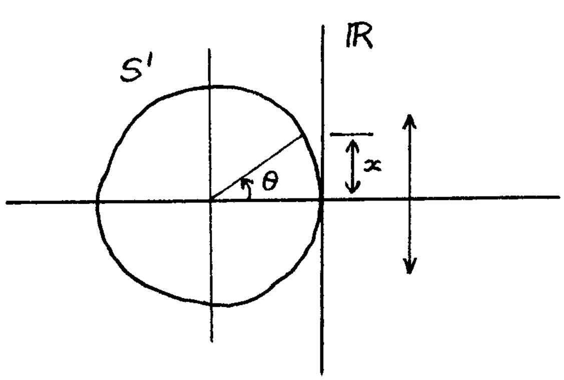
Note the link between $J : C_1^s \to C_2^s$ and the exponential map $\exp : \mathfrak{o}(2) \to SO(2)$. Lo and Willsky state that J^{-1} is given by

$$(J^{-1}A)(t) = \int_0^t [A^T(s)\, dA(s)]_{12}.$$

Given a continuous $\mathcal{A}$-measurable random process $Y : \Omega \to C_2^s$, there corresponds a unique $\mathcal{A}$-measurable random process $y : \Omega \to C_1^s$, such that $y(t) = (J_y)(t)$, and vice versa. We speak of process pairs (y, Y), where y is a random process on $\mathbb{R}$ and Y is the corresponding random process on S^1.

Brownian Motion on S^1

Brownian motion on the real line is used to define Brownian motion on the circle.



Let the line more up and down under Brownian motion. This rolls the circle, inducing Brownian motion on it. This is formulated precisely in Itô and McKean[1]

The probability density satisfies the heat equation on the circle

$$\frac{\partial p_\theta}{dt} - \frac{1}{2}\frac{\partial^2 p_\theta}{\partial \xi^2} = 0$$

subject to

$$p_\theta(\xi, t) = p_\theta(\xi + 2\pi, t); \quad p_\theta(\xi, 0) = \delta(\xi - \theta_0).$$

The solution is called the "folded normal density":

$$p_\theta(\xi, t) = \frac{1}{\sqrt{2\pi t}} \sum_{n=-\infty}^{\infty} e^{-[(\xi + 2n\pi - \theta_0)^2/2t]}.$$

In general, given a process pair (y, Y), the density for Y is given by "folding" the density for y around the circle.

<u>Observation and Message Processes</u>

Let (z, Z) be an observation of the message (x, X). By definition, the process $z : \Omega \to c_1^s$ satisfies the Itô differential equation

$$dz(t) = h(x(t), t)dt + r^{1/2}(t)dw(t); \quad z(0) = 0;$$

where w is a Brownian motion independent of x, h is Borel measurable and r is positive and Borel measurable. This models sensor dynamics.

Z is related to z via $Z = J(z)$, and applying the Itô differential rule gives the matrix Itô differential equation

$$dZ(t) = \left[-\frac{1}{2}r(t)Idt + Rdz(t)\right] Z(t); \quad Z(0) = I.$$

$Z(t)$ can be written

$$Z(t) = \exp\left[R\int_0^t r^{1/2}(s)dw(s)\right] \exp\left[R\int_0^t h(x(s), s)ds\right].$$

The message process x satisfies the Itô differential equation

$$dx(t) = a(x(t), t)dt + b^{1/2}(x(t), t)dv(t),$$

[1] Itô, K. and McKean, H.P., "Diffusion Processes and Their Sample Paths.", New York, Academic, 1965.

where v is a Brownian motion, a is Borel measurable, and b is positive and Borel measurable. This is the message signal model.

Lo and Willsky point out that even though x is a Markov process, X need not be. X satisfies

$$dX(t) = \left[-\frac{1}{2}b(x(t),t)Idt + Rdx(t)\right] X(t).$$

Depending on the application, we may want to estimate x (e.g. frequency demodulation) or X (e.g. angular orientation).

The sensor can observe either Z (e.g. angular position), or dZ (e.g. angular change). Note that the input to the sensor is x, not X. Thus the observations yield information about X (e.g. angular position) and x (e.g. total angle swept).

Conditional Probability Distributions

Define $z^t = \{z(\tau) : \tau \in [0,t]\}$ and $Z^t = \{Z(\tau) : \tau \in [0,t]\}$. The σ-algebra $\mathcal{A}_Z^t \subset \mathcal{A}$ generated by z^t coincides with the σ-algebra generated by Z^t, since J is bijective. The information contained in z^t is the same as that in Z^t. So conditioning on z^t is equivalent to conditioning on Z^t. In general, the σ-algebra $\mathcal{A}_{x(t)}$ generated by $x(t)$ differs from the σ-algebra $\mathcal{A}_{X(t)}$ generated by $X(t)$.

Lo and Willsky deduce the stochastic partial differential equations for the conditional densities $P_x(x,t|Z^t)$ and $P_\theta(\theta,t|Z^t)$ for the filtering problem:

$$dp_x(x,t|Z^t) = A^* p_x(x,t|Z^t)dt + \frac{(h(x,t) - \hat{h}_t)}{r(t)}([Z^T(t)dZ(t)]_{12} - \hat{h}_t dt)p_x(x,t|Z^t)$$

where

$$\begin{aligned} A^*(g) &= -\frac{\partial(ag)}{\partial x} + \frac{1}{2}\frac{\partial^2(bg)}{\partial x^2} \\ \hat{h}_t &= E(h(x,t)|Z^t) \\ dp_\theta(\theta,t|Z^t) &= \sum_{n=-\infty}^{\infty} dp_x(\theta + 2n\pi, t|Z^t). \end{aligned}$$

Optimal Filtering Equations

Explicit equations for our optimal non-linear estimation problem can be obtained when the model equations are linear. Message process

$$\begin{aligned} dx(t) &= a(t)x(t)dt + b^{1/2}(t)dv(t); \quad x(0) = 0; \\ dX(t) &= \left[-\frac{1}{2}b(t)Idt + Rdx(t)\right] X(t); \quad X(0) = I. \end{aligned}$$

Observation process

$$\begin{aligned} dz(t) &= c(t)x(t)dt + r^{1/2}(t)dw(t), \;\; z(0) = 0; \\ dZ(t) &= \left[-\frac{1}{2}r(t)I + c(t)x(t)R\right] Z(t)dt + r^{1/2}(t)RZ(t)dw(t), \;\; z(0) = I. \end{aligned}$$

The optimal filtering estimates $\hat{x}(t|t)$ and $\hat{X}(t|t)$ are $\mathcal{B}_1^s$ - and $\mathcal{B}_2^s$ - measurable functions of Z^t, satisfying

$$\begin{aligned} E(\phi_1(X(t), \hat{X}(t|t))|Z^t) &\leq E(\phi_1(x(t), M)|Z^t) \\ E(\phi_2(x(t) - \hat{x}(t|t))|Z^t) &\leq E(\phi_2(x(t) - \mu)|Z^t), \end{aligned}$$

where $\phi_1 : S^1 \to I\!R^1$ and $\phi_2 : I\!R^1 \to I\!R^1$ are suitable error functions, for all $\mathcal{A}_Z^t$-measurable $M : \Omega \to SO(2)$, and all $\mathcal{A}_Z^t$-measurable $\mu : \Omega \to I\!R$.

Standard Kalman-Bucy linear filtering theory can be applied to give $\hat{x}(t|t)$ in terms of $dz(t)$. The conditional density is

$$p_x(x, t|z^t) = \frac{1}{\sqrt{2\pi P(t)}} \exp\left[-\frac{(x - \hat{x}_{t|t})^2}{2P(t)}\right],$$

with filtering equations, expressed in terms of Z,

$$\begin{aligned} d\hat{x}(t|t) &= a(t)\hat{x}(t|t)dt \\ &\quad + P(t)c(t)r^{-1}(t)\left([Z^T(t)dZ(t)]_{12} - c(t)\hat{x}(t|t)dt\right), \; \hat{x}(0|0) = 0; \\ \dot{P}(t) &= 2a(t)P(t) - c^2(t)r^{-1}(t)P^2(t) + b(t), \;\; P(0) = 0. \end{aligned}$$

Lo and Willsky note that the conditional expectation $E(x(t)|\mathcal{A}_Z^t)$ is both a $\mathcal{B}_1^t$-measurable function of Z^t and a $\mathcal{B}_2^t$-measurable function of Z^t.

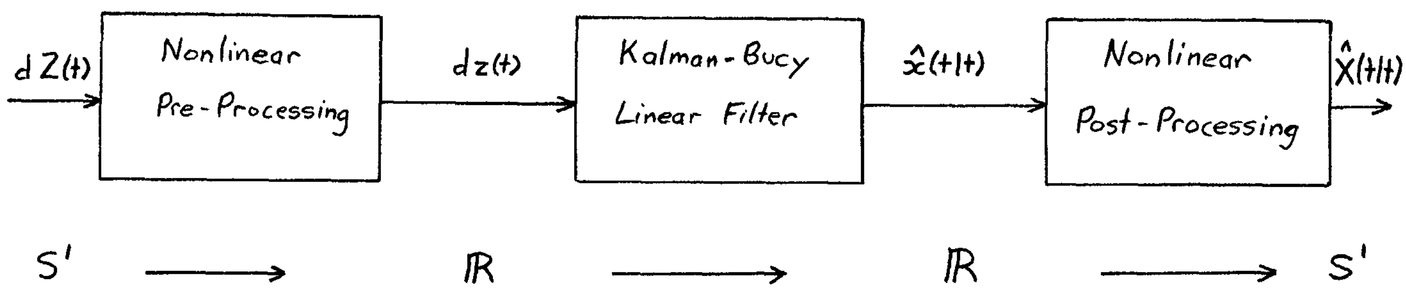
Estimates on the circle are given by

$$\hat{\theta}_x(t|t) = \hat{x}(t|t) \mod 2\pi, \quad \text{or} \quad \hat{X}(t|t) = \exp[\hat{x}(t|t)R].$$

The optimal estimate is not the conditional expectation $E(X(t)|Z^t)$, but is given by the equations

$$\begin{aligned} d\hat{X}(t|t) &= -\frac{1}{2}p^2(t)c^2(t)r^{-1}(t)\hat{X}(t|t)dt + \{a(t) - P(t)c^2(t)r^{-1}(t) \\ &\quad \left[\int_0^t (\exp \int_s^t (a(\tau) - P(\tau)c^2(\tau)r^{-1}(\tau))d\tau)P(s)c(s)r^{-1}(s)[Z^T(s)dZ(s)]_{12}dt\right] \\ &\quad + P(t)c(t)r^{-1}(t)[Z^T(t)dZ(t)]_{12}\}R\hat{X}(t|t); \quad \hat{X}(0|0) = I \end{aligned}$$

We can summarise optimal filtering on the circle as follows.
Measurements dZ are processed by a non-linear transformation $dz = [Z^T dZ]_{12}$, essentially J^{-1}. We are going from the Lie group S^1 to its Lie algebra $I\!R$. Then a standard linear Kalman-Bury filter processes dz to yield $\hat{x}(t|t)$. Thus the filtering takes place in the Lie algebra. Then $\hat{x}(t|t)$ is transformed by J to give the desired estimate $\hat{X}(t|t)$. Here we're going from the Lie algebra $I\!R$ back to the Lie group S^1.



Extensions

The above method can be extended directly to Abelian Lie groups, $I\!R^n \times T^m$. Just transform to the Lie algebra $I\!R^{n+m}$, use a linear filter, and transform back again.
There have been efforts to develop optimal estimation filters for non-abelian Lie groups. The bracket structure then plays a curcial role.

References

[1] Lo, J.T. and Willsky, A.S., *"Estimation for Rotational Processes with One Degree of Freedom"*, Parts I, II, III; IEEE Trans. Aut. Control, Vol, AC-20, No. 1, Feb. 1975.

[2] Marcus, S.I., *"Algebraic and Geometric Methods in Non-linear Filtering."*, SIAM J. Control and Optimization, Vol. 22, No. 6, Nov. 1984.

[3] Warner, F.W., *"Foundations of Differentiable Manifolds and Lie Groups."*, Springer-Verlag, 1983.

[4] Spivak, M., *"A Comprehensive Introduction to Differential Geometry"*, Volume 1, Second Edition, Publish or Perish, 1979.

[5] Grove, K., *"Differential Geometry"*, Lecture notes for MATH 740, Mathematics Department, University of Maryland, Spring 1984.