ODEs, Integral Flows and Frobenius Theorem

Shrivathsa Pandelu

November 20, 2022

Contents

1 Solving a system of equations

Let D be some open set in $\mathbb{R}^{n+1}$ and let $f_1, \ldots, f_n$ be continuous functions defined on D. We seek to obtain differentiable functions $x_i(t), 1 \le i \le n$ such that

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = f_i(x_1, \ldots, x_n, t), 1 \le i \le n$$

on some interval (t_1, t_2).

Following [1] we shall denote vectors by capital letters, scalars by small letters and the norm of a vector is the usual Pythagorean norm. A vector function of t is a function $X(t) = (x_1(t), \ldots, x_n(t))$ and as usual, it is continuous, differentiable etc. if each of the components is. Denote by F the collection of functions $F = (f_1, \ldots, f_n)$, then we wish to solve

$$\frac{\mathrm{d}X}{\mathrm{d}t} = F(X, t) \tag{1}$$

on some interval (t_1, t_2). Let $X_0 = (x_{10}, \ldots, x_{n0}), t_0$ be fixed in the given domain D, and suppose F is defined on a region $R = |X - X_0| \le a, |t - t_0| \le b$, then we would also like to have a solution satisfying $X(t_0) = X_0$.

Definition 1. *An approximate solution to 1 with error ϵ is a continuous admissible function (i.e., the image should land where F is defined) X defined on some interval around t_0 which is piecewise differentiable such that $X(t_0) = X_0$ and $|X'(t) - F(X(t), t)| \le \epsilon$ wherever X is differentiable.*

Since R is compact, F is continuous it is bounded by some constant M.

Theorem 1. (*Existence of approximate solutions*) *With a setup as above, given an $\epsilon > 0$ we can construct an approximate solution with error ϵ over the interval $|t - t_0| \leq h = \min(a, b/M)$.*

Proof. The idea is to form an approximate solution as a union of lines (this is somewhat related to how integral of a function is defined using step functions). The constriants on h arise from requiring the lines to lie in R.

Given $\epsilon > 0$ by uniform continuity of F on R, there's a $\delta > 0$ such that if $|X_1 - X_2| < \delta, |t_1 - t_2| < \delta$, then $|F(X_1, t_1) - F(X_2, t_2)| < \epsilon$. We first define our approximate solution to the right of t_0. Obtain numbers $t_0 < t_1 < \cdots < t_m = t_0 + h$ such that

$$|t_{i+1} - t_i| < \min(\delta, \delta/M), i = 0, \ldots, m - 1.$$

The approximate solution is defined recursively as

$$X_i(t) = X_{i-1}(t_i) + F(X_{i-1}(t_i), t_i)(t - t_i), t \in [t_i, t_{i+1}], 0 \leq i \leq m - 1$$

with $X_{-1}(t_0) = X_0$. The union of these X_is form a continuous function which is piecewise differentiable. We need to verify that this is admissible. By the choice of h, there is a cone starting at X_0 that is contained in R whose boundary is the union of lines with slope M going to the right of t_0 starting from X_0. Each of the line segments X_i is contained in that cone and recursively we see that this solution is admissible. More directly, the end points of the line segments are within $M\delta$ of each other and adding them up gives a distance of $M\delta \times h/\delta \leq b$.

Next, by the choice of δ, for $t_i \leq t \leq t_{i+1}$ we have

$$|X_i(t) - X_i(t_i)| < \delta, t - t_i < \delta$$

therefore $|F(X_i(t), t) - X_i'(t)| = |F(X_i(t), t) - F(X_i(t_i), t_i)| < \epsilon$. Thus we have constructed an approximate solution to the right of t_0. In a similar manner, but going to the left we get an apprximate solution with error ϵ defined on $|t - t_0| \leq h$. □

In order to guarantee exact solutions we will need additional conditions on F, namely Lipschitz continuity with respect to X. This is not too restrictive a condition because on usually deals with smooth F, and being defined on compact sets (or if we are just interested in existence of solutions we may restrict our attention to compact sets using local compactness) it will be Lipschitz.

2 Lipschitz continuity and properties

Definition 2. *A (continuous) vector function $F(X, t)$ defined on some open $D \subseteq \mathbb{R}^{n+1}$ satisfies the Lipschitz condition with respect to X there is a constant k such that $|F(X_1, t) - F(X_2, t)| \leq k|X_1 - X_2|, (X_1, t), (X_2, t) \in D$. The same defintion applies to other domains of definition and when F is scalar.*

Lemma 1. *Let R be a region convex in X, i.e., for each t_0, the subset $t = t_0$ of $\mathbb{R}^n$ is convex (or empty), and suppose the scalar function f has partial derivatives $\frac{\partial f}{\partial x_i}$ defined on R and bounded by some N, then f satisfies the Lipschitz condition on X with constant nN.*

Proof. Because of convexity, we may apply the mean value theorem. For a given $(X_1, t), (X_2, t) \in R$ we have

$$|f(X_1, t) - f(X_2, t)| = |\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_{i1} - x_{i2})| \leq N \sum_{i=1}^{n} |x_{i1} - x_{i2}| \leq nN|X_1 - X_2|$$

where x_{i1}, x_{i2} are the ith components of X_1, X_2 respectively. We can in fact, using Cauchy-Schwartz inequality obtain the better constant $\sqrt{n}N$. □

Lemma 2. *Let R be a region embedded in a larger domain D such that the boundaries of the two have a minimum distance of some $\delta > 0$. Suppose scalar functions $f, \partial f/\partial x_i$ are defined and bounded by M, N respectively on D, then f satisfies the Lipschitz condition on X in R with constant* $\max(2M/\delta, nN)$.

Proof. Let $(X_1, t), (X_2, t) \in R$. If $|X_1 - X_2| \geq \delta$, then

$$\frac{|f(X_1, t) - f(X_2, t)|}{|X_1 - X_2|} \leq \frac{2M}{\delta}.$$

If $|X_1 - X_2| < \delta$, then the line segment between (X_1, t) and (X_2, t) must lie in D. For otherwise (by connectedness of the line), it must include points from the boundaries of R, D which would contradict the definition of δ. Now we may apply the argument of the previous lemma to obtain the constant nN. □

Lemma 3. *A vector function F satisfies the Lipschitz condition on X if and only if each of its components satisfy the Lipschitz condition on X.*

Proof. The proof is quite simple and left to the reader. □

Lemma 4. *If a vector function $X(t)$ is defined and differentiable on some interval (t_1, t_2), then the scalar function $|X(t)|$ is differentiable on the open set $X(t) \neq 0$ and we have*

$$\left|\frac{\mathrm{d}|X|}{\mathrm{d}t}\right| \leq \left|\frac{\mathrm{d}X}{\mathrm{d}t}\right|.$$

Proof. On the open set where $X \neq 0$, we have by chain rule

$$\frac{\mathrm{d}|X|}{\mathrm{d}t} = \frac{\sum x_i \frac{\mathrm{d}x_i}{\mathrm{d}t}}{|X|}.$$

Furthermore,

$$\left|\frac{|X(t+h)| - |X(t)|}{h}\right| \leq \left|\frac{X(t+h) - X(t)}{h}\right|$$

and since the limit exists, the same inequality holds for the derivatives. □

Theorem 2. (*Fundamental inequality*) *Let $F(X, t)$ be continuous in some region R and satisfy the Lipschitz condition on X with constant k. Let $X_1(t), X_2(t)$ be two approximate solutions over an interval $|t - t_0| \leq b$ with errors ϵ_1, ϵ_2 respectively. Set*

$$P(t) = X_1(t) - X_2(t); p(t) = |P(t)|; \epsilon = \epsilon_1 + \epsilon_2.$$

Then

$$|p(t)| \leq e^{k|t-t_0|}|p(t_0)| + \frac{\epsilon}{k}(e^{k|t-t_0|} - 1)$$

for $|t - t_0| \leq b$.

Proof. P is differentiable wherever both X_1, X_2 are differentiable, hence it is differentiable at all but finitely many points. Wherever P is differentiable we by the previous lemma

$$|p'(t)| \leq |P'(t)|$$

and by definition of an approximate solution

$$|P'(t)| \leq \epsilon + |F(X_1(t), t) - F(X_2(t), t)| \leq \epsilon + k|X_1(t) - X_2(t)| = \epsilon + k|p(t)|.$$

Therefore, we have a continuous function $p(t)$ on $|t - t_0| \leq b$ such that

$$|p'(t)| \leq k|p(t)| + \epsilon$$

wherever p is differentiable. Since p is differentiable at all but finitely many points, we can meaningfully talk about integrating p' and the fundamental theorem of calculus applies.

Suppose p is always positive, then to the right of t_0,

$$e^{-kt}(p'(t) - kp(t)) \leq e^{-kt}\epsilon$$

and integrating from t_0 to $t > t_0$ we get the desired inequality. To the left of t_0, use $-p' \leq kp + \epsilon$ instead.

If p is identically 0, then the inequality is obvious. Now, suppose $p(t) \neq 0$ for some $t > t_0$ and that there is a $t_0 < t' < t$ such that $p(t') = 0$. Using continuity, one obtains $t_0 < t_1 < t$ such that $p(t_1) = 0$ and p is non zero on (t_1, t). Now apply the previous case to obtain

$$|p(t)| \leq \frac{\epsilon}{k}(e^{k(t-t_1)} - 1)$$

which is better than the bound in the statement. One does a similar thing for points to the left of t_0. For p being negative, simply use $-p$. □

3 Existence and Uniqueness

From the fundamental inequality, we have the following

Theorem 3. (*Uniqueness*) *Let $F(X, t)$ be continuous and satisfy the Lipschitz condition on X in some neighbourhood of a fixed point (X_0, t_0). Then there can exist at most one solution of $dX/dt = F(X(t), t)$ such that $X(t_0) = X_0$.*

Proof. While the fundamental inequality applied to (closed) intervals around t_0, we can adapt the same proof here as well, since we are dealing with an open neighbourhood of (X_0, t_0). □

Theorem 4. (*Existence*) *Let $F(X, t)$ be continuous and satisfy a Lipschitz condition on X in a region*

$$|X - X_0| \leq b, |t - t_0| \leq a.$$

Let M be the upper bound of $|F|$ in R. Then there exists a solution $X(t)$ of $dX/dt = F(X, t)$ defined over the interval $|t - t_0| \leq h = \min(a, b/M)$.

Proof. The idea is to take the uniform limit of approximate solutions. Let ϵ_n be a sequence of positive numbers decreasing to 0 and let $X_n(t)$ be approximate solutions with error ϵ_n defined over $|t-t_0| \leq h$ passing through X_0 at t_0.

Step 1

Applying the fundamental inequality,

$$|X_n(t) - X_{n+m}(t)| \leq \frac{\epsilon_n + \epsilon_{n+m}}{k}(e^{k|t-t_0|} - 1) \leq \frac{2\epsilon_n}{k}(e^{kh} - 1).$$

Therefore, these X_n converge uniformly to some function X. By uniform convergence, X is continuous and $X(t_0) = X_0$. We need to show that X is differentiable with the desired derivative. We first show that the approximate derivatives of $F(X_n, t)$ converge uniformly to $F(X, t)$.

Step 2

Let R be the region $|X - X_0| \leq Mh, |t - t_0| \leq h$. Given $\epsilon > 0$ by uniform continuity of F, there is a $\delta > 0$ such that whenever $(X_1, t_1), (X_2, t_2) \in R$ and $|X_1 - X_2| < \delta, |t_1 - t_2| < \delta$ we have $|F(X_1, t_1) - F(X_2, t_2)| < \epsilon$.

For this δ, there is an N such that for $n \geq N, |X_n(t) - X(t)| < \delta, |t - t_0| \leq h$. Therefore, for $n \geq N$

$$|F(X, t) - F(X_n, t)| < \epsilon, n \geq N, |t - t_0| \leq h.$$

Thus, $F(X_n, t)$ converges uniformly to $F(X, t)$ on $|t - t_0| \leq h$, hence the integrals $\int_{t_0}^t F(X_n, t)dt$ converge to $\int_{t_0}^t F(X, t)$ on $|t - t_0| \leq h$.

Step 3

We have shown above that X_n converge uniformly to X, their derivatives converge to $F(X,t)$ uniformly. Given any $t \in |t - t_0| \leq h$ we have

$$\left| \int_{t_0}^{t} [\frac{\mathrm{d}X_n}{\mathrm{d}t} - F(X_n, t)]dt \right| \leq \epsilon_n h$$

therefore, applying the fundamental theorem of calculus (recall that X_n are differentiable at all but finitely many points)

$$|X_n(t) - X_n(t_0) - \int_{t_0}^{t} F(X_n, t)dt| \leq \epsilon_n h$$

We can take the limit of the left side and by uniform convergence, we have

$$|X(t) - X_0 - \int_{t_0}^{t} F(X(t), t)dt| = 0$$

which means that X is differentiable with derivative $F(X,t)$ as required. □

Now that we have shown the existence and uniqueness of solutions, we observe that our approximate solutions as defined above are indeed approximations of the exact solution.

Theorem 5. *Let X be an exact solution of 1 under the assumptions of the previous theorem and $\tilde{X}$ be an approximate solution with error ϵ for $|t - t_0| \leq h$ such that $\tilde{X}(t_0) = X_0$. Then there is a constant N independent of ϵ such that $|\tilde{X}(t) - X(t)| \leq \epsilon N, |t - t_0| \leq h$.*

Proof. Applying the fundamental inequality to $\tilde{X}, X$,

$$|\tilde{X}(t) - X(t)| \leq \frac{\epsilon}{k}(e^{kh} - 1); \text{ i.e., } N = \frac{e^{kh} - 1}{k}.$$ □

Without the Lipschitz condition, the solution need not be unique: consider $dy/dx = y^{1/3}$ with $f(x,y) = y^{1/3}$ which is continuous. There are two solutions passing through $(0,0)$, namely $y \equiv 0$ and

$$y(x) = \begin{cases} (\frac{2}{3}x)^{3/2} & x \geq 0 \\ 0 & x \leq 0 \end{cases}$$

However, the Lipschitz condition is not required for the existence of solutions. Once can use polynomial approximations of F using the Weierstrass approximation theorem together with the Arzela-Ascoli theorem. Essentially, if P_n are polynomials uniformly converging to F, and if X_n are exact solutions to $dX/dt = P_n$, then one can show that there is a subsequence of X_n converging uniformly to some X and that this X is the required solution. For a proof see [1].

Before moving to the next part, it is good to mention that although above we arrived at an exact solution using Cauchy-Euler approximations, this is not the only way to arrive at a solution. One such way is Picard's theorem which uses a fixed point theorem (contraction theorem). This is similar to Newton's method for finding roots in that one starts with an arbitrary function and generates successive functions which eventually arrive at a solution. Furthermore, there are refinements of the Cauchy-Euler construction, see [1].

4 Smoothness of solutions

In the previous section our solution was essentially a curve passing through a given point (X_0, t_0). Now we look at how it the family of these curves depends on the initial condition X_0.

First we observe that if F is of class $\mathcal{C}^k$, then a solution $X(t)$ defined around some interval around t_0 is of class $\mathcal{C}^{k+1}$. When considering vector fields over manifolds, as we shall later, we can first restrict our attention to some coordinate chart, obtain a smooth solution over an interval and then try to patch up our solutions (or simply deal with maximal intervals where a smooth solution exists).

Theorem 6. *If $F(X,t)$ is continuous and satisfies a Lipschitz condition on X in $|X - \tilde{X}_0| \leq b, |t - t_0| \leq a$, for some fixed $(\tilde{X}_0, t_0)$, then there exists a unique solution $X(X_0, t)$ of*

$$X' = F(X,t), X(t_0) = X_0$$

in the region

$$|X_0 - \tilde{X}_0| \leq b/2, |t - t_0| \leq h', |X - X_0| \leq Mh'$$

where $h' = \min(a, (b/2M))$, M being the bound of F on $|X - \tilde{X}_0| \leq b, |t - t_0| \leq a$.

Proof. The given region contains around (X_0, t_0) the rectangle $|X - X_0| \leq b/2, |t - t_0| \leq a$ for $|\tilde{X}_0 - X_0| \leq b/2$. From the existence theorem, a solution to $X' = F(X,t)$ exists in the interval as mentioned. Note that h' is no less than half of the original h. □

What this means is that there is a smaller neighbourhood of a given $(\tilde{X}_0, t_0)$ where we have a family of solutions passing through every point (X_0, t_0). Note that we are not changing t_0. In the context of vector fields on manifolds, we may take $t_0 = 0$ and the theorem says that there is a flow of a neighbourhood of $(\tilde{X}_0, 0)$, which gives us, in a sense, a time parameterised family of diffeomorphisms which at $t = 0$ gives us the identity morphism and such that each point moves along the vector field determined by F.

Theorem 7. *If in some open neighbourhood of $(\tilde{X}_0, t_0), F(X,t)$ is continuous and satisfies a Lipschitz condition on X for some k, and if a solution $X(X_0, t)$ of*

$$X' = F(X,t), X(t_0) = X_0$$

exists in some rectangle $R, |t - t_0| \leq h, |X_0 - \tilde{X}_0| \leq l$, then $X(X_0, t)$ is continuous in X_0, t simultaneously in R.

Here we are considering the following function. Given $(X_0, t) \in R$, its image is the value of the curve passing through (X_0, t_0) at time t. The first coordinate gives us a curve, and the second gives us a point on this curve.

Proof. This is a simple application of the fundamental inequality. Fix a point $(X_0^{(1)}, t_1)$. For any $X_0^{(2)}$, we have by the fundamental inequality

$$|X(X_0^{(1)}, t) - X(X_0^{(2)}, t)| \leq |X_0^{(1)} - X_0^{(2)}| e^{kh}, |t - t_0| \leq h.$$

Now, given $\epsilon > 0$, choose a $\delta > 0$ so that

$$|t - t_1| < \delta \implies |X(X_0^{(1)}, t) - X(X_0^{(1)}, t_1)| < \epsilon$$

Combining the two inequalities and applying the triangle inequality, it is clear that $X(X_0, t)$ is continuous at $(X_0^{(1)}, t_1)$. □

The fundamental inequality shows that the solution is uniformly continuous with respect to the intial value, this means that two curves are never farther apart than their initial points, upto a constant.

Lemma 5. *Let C be a compact subset of $\mathbb{R}^{n+m}, n \geq 1, m \geq 0$ and let f be a scalar function defined on C. Write points of $\mathbb{R}^{n+m}$ as $(x_1, \ldots, x_n, y)$ or (x, y). Suppose f extends to a neighbourhood U of C and has continuous partial derivatives $\partial f / \partial x_i$ in U, then*

$$\frac{f(x+h, y) - f(x,y) - \sum_{i=1}^{n} \partial_i f(x,y) h_i}{|h|}$$

approaches 0 as $h \to 0$ uniformly over C.

Proof. In the proof below, we use $|x|$ to denote the absolute value of real numbers and magnitudes of vectors. For each $p \in C$ there is an $r_p > 0$ such that $B^{n+m}(p, 2r_p) \subseteq U$. Cover C by $\{B^{n+m}(p, r_p)\}_{p \in C}$ and obtain a finite subcover $B^{n+m}(p_1, r_1), \ldots, B^{n+m}(p_k, r_k)$ and let $r = \min\{r_1, \ldots, r_k\}$. Given $p \in C$, it is clear that $B^{n+m}(p, r) \subset U$. So we may define the error function

$$
\begin{aligned}
&E \colon C \times B(0,r) \to \mathbb{R} \\
&((x,y),h) \mapsto \begin{cases} \frac{f(x+h,y)-f(x,y)-\sum_{i=1}^{n} \partial_i f(x,y) h_i}{|h|} & h \neq 0 \\ 0 & h = 0 \end{cases}
\end{aligned}
$$

where $B(0,r)$ is the ball of radius r around the origin in $\mathbb{R}^n$. We wish to show that given any $\epsilon > 0$ there is a $\delta > 0$ such that for $|h| < \delta, |E((x,y),h)| < \epsilon \,\forall\, (x,y) \in C$. Fix an $\epsilon > 0$, we know that for each $p = (x,y) \in C$ there is a $\delta_p > 0$ such that the inequality holds at p. We first show that the same δ_p works on a neighbourhood of (x,y).

For two points $p = (x,y), q = (x', y')$ and an $h \in B(0,r), h \neq 0$, write $p_h = (x+h, y), q_h = (x'+h, y')$, then

$$
\begin{aligned}
|E(q,h)||h| &\leq |E(p,h)||h| + |f(q_h) - f(q) - f(p_h) + f(p)| + \sum_{i=1}^{n} |\partial_i f(p) - \partial_i f(q)||h_i| \\
&\leq |E(p,h)||h| + |\sum_{i=1}^{n} \partial_i f(z', y') h_i - \sum_{i=1}^{n} \partial_i f(z, y) h_i| + \sum_{i=1}^{n} |\partial_i f(p) - \partial_i f(q)||h_i| \\
&\leq |E(p,h)||h| + \sum_{i=1}^{n} |\partial_i f(z', y') - \partial_i f(z,y)||h_i| + \sum_{i=1}^{n} |\partial_i f(p) - \partial_i f(q)||h_i|
\end{aligned}
$$

where z is some point on the line between $x, x+h$ and z' on the line between $x', x'+h$ assuming that these lines are in U. This assumption is justified if the above calculation is done in some suitable ball around (x,y).

Now fix $p_0 \in C$ and $\epsilon > 0$. There is a δ_0 such that for $|h| < \delta_0, E(p_0, h) < \epsilon$. Using uniform continuity of f and its derivatives on C obtain a $\delta > 0$ such that for two points within δ the values of f and its partial derivatives are within ϵ. We can take $4\delta_0 < \delta < r$. Let q be within $\delta/4$ of p_0, then for $|h| < \delta_0$, from the calculation above we have

$$
|E(q,h)|h| \leq |E(p_0,h)||h| + n\epsilon|h| + n\epsilon|h| \leq (2n+1)\epsilon|h|.
$$

It follows that for a given $\epsilon > 0, p_0 \in C$, there is are $\delta, \delta_0 > 0$ such that for any q within δ of p_0, and $|h| < \delta_0$, we have $|E(q,h)| < \epsilon$. Cover C with such open balls and take a finite subcover. Taking the minimum of all the δ_0s, it follows that $E(p,h)$ converges uniformly to 0 as $h \to 0$ on C. $\square$

Remark. Note that the $\mathbb{R}^m$ factor can be replaced by any other metric space, the presence of the second coordinate doesn't affect the proof in anyway. We can apply mean value theorem because we are dealing with a scalar function, and fixing the second coordinate gives us a $\mathcal{C}^1$ function on some open subset of $\mathbb{R}^n$.

Theorem 8. *Let $F(X,t)$ be continuous and satisfy a Lipschitz condition on X in some region $R : |X - \tilde{X}_0| \leq b, |t - t_0| \leq a$ with upper bound M. Suppose a solution exists on R for all initial points. Write*

$$
\begin{aligned}
&F = (f_1, \ldots, f_n) \\
&X(X_0, t) = (x_1(X_0,t), \ldots, x_n(X_0,t)) \\
&X_0 = (x_{10}, \ldots, x_{n0})
\end{aligned}
$$

Suppose all partial derivative $\frac{\partial f_i(X,t)}{\partial x_j}, 1 \leq i,j \leq n$ exist and are continuous simultaneously in X, t on R. Then $\frac{\partial x_i(X_0,t)}{\partial x_{j0}}, 1 \leq i,j \leq n$ exist and are continuous in X_0, t simultaneously over R.

Note here that we have assumed a solution exists with any initial point in R, usually we would have to shrink R and then apply the theorem.

Proof. By shrinking R if necessary, we may also assume that F and its derivatives exist on a neighbourhood of R. We will show that $\partial x_i(X_0, t)/\partial x_{10}$ exist for all i, the other derivatives are similar. Take two points $X_0, X_0 + (\Delta x_{10}, 0, \ldots, 0)$. With these points as initial conditions, we have solutions

$$(x_1(X_0, t), \ldots, x_n(X_0, t)) \text{ and } (x_1(X_0, t) + \Delta x_1(X_0, t), \ldots, x_n(X_0, t) + \Delta x_n(X_0, t))$$

respectively. Note that these functions depend on the initial point and time. By the fundamental inequality, we have

$$\Delta = [\Delta x_1(X_0, t)^2 + \cdots + \Delta x_n(X_0, t)^2]^{1/2} \leq e^{kb}|\Delta x_{10}|.$$

Set

$$p_i(X_0, t) = \frac{\Delta x_i(X_0, t)}{\Delta x_{10}}$$

We must show that $\lim_{\Delta x_{10} \to 0} p_i(X_0, t)$ exists. The idea is as follows. Since p_i is a constant times the difference of solutions, differentiating it would give us a relation in terms of f_i. We will show that p_i is an approximate solution to another system of differential equations and that these p_i converge to the solution of that system.

To this end, differentiating Δx_i gives

$$\begin{aligned}\frac{\partial \Delta x_i}{\partial t}(X_0, t) &= f_i(x_1 + \Delta x_1, \ldots, x_n + \Delta x_n, t) - f_i(x_1, \ldots, x_n, t) \\ &= \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(x_1, \ldots, x_n, t)\Delta x_j(X_0, t) + \eta_i(x_1, \ldots, x_n, t)\Delta\end{aligned}$$

where η_i is the error function. A couple of things worth mentioning. Our hypothesis is that each f_i has partial derivatives in x_j which are continuous as functions of X, t, since R is compact, f satisfies the hypothesis of the lemma. Secondly, note that all functions on the right are functions of the initial point and time.

By the lemma above, $\eta_i \to 0$ uniformly on R as $\Delta \to 0$. Here although the x_is depend on the initial value and t, we know that as long as Δ is small, η_i is also small, regardless of what $(x_1, \ldots, x_n, t)$ is. However, Δ depends on the initial point and t. This is where the fundamental inequality helps because we can bound Δ uniformly by Δ_{10}.

So, as $\Delta_{10} \to 0, \Delta \to 0$. So, independent of where $(x_1, \ldots, x_n)$ is (which depends on X_0, t), $\eta_i(x_1, \ldots, x_n, t)$ goes to 0 as well. Dividing by Δx_{10} we get

$$\frac{\mathrm{d}p_i}{\mathrm{d}t}(X_0, t) = \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(x_1, \ldots, x_n, t)p_j + \eta_i \frac{\Delta}{\Delta x_{10}}.$$

So, if we fix an initial point $X_0, \Delta x_{10}$, then p_i (now a function of t alone) are an approximate solution of the system

$$\frac{\mathrm{d}q_i}{\mathrm{d}t} = \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(x_1(t), \ldots, x_n(t), t)q_j \tag{2}$$

where now x_i are treated as a function of time, with initial condition

$$q_i(t_0) = \begin{cases} 1 & i = 1 \\ 0 & i \neq 1 \end{cases}$$

It is easy to see that the derivatives of q_i, which are linear in q with coefficients depending on t, are continuous in q_i, t and satisfy the Lipschitz condition on q_is because the derivatives of f are continuous , hence bounded, on R. Therefore a solution exists with the given initial conditions, moreover it exists on $|t - t_0| \leq a$ because it is a linear system.

Moreover, $p_i(X_0, t)$ satisfy the same system with error less than

$$n \max_i (\eta_i(x_1(X_0,t), \ldots, x_n(X_0,t), t) \frac{\Delta}{\Delta x_{10}})$$

which goes to 0 uniformly, independent of X_0, t, hence can be made arbitrarily small.

Having fixed X_0, because η_is go uniformly to zero over R as $\Delta x_{10} \to 0$ and $\Delta/\Delta x_{10}$ is bounded uniformly on R, it follows that as functions of t on $|t - t_0| \leq a$, the approximate solutions p_i converge to a solution of 2. Therefore, for a fixed X_0, the limit

$$\lim_{\Delta x_{10} \to 0} p_i(t)$$

exists, which means that the $(x_1, \ldots, x_n)$ are differentiable as functions of x_{10}. Note importantly that the limit of these p_is exist *uniformly* on R, i.e., the derivative quotient converges at the same rate at all points of R. We will prove the continuity of these derivatives later. □

We spend some time understanding the system 2. On R, let $X(X_0, t) = (x_1(X_0, t), \ldots, x_n(X_0, t))$ be a solution of $dX/dt = F(X, t)$. Then, the partial derivative of x_i with respect to the first coordinate x_{10} of X_0 satsify

$$\frac{\mathrm{d}q_i}{\mathrm{d}t}(X_0, t) = \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(x_1(X_0,t), \ldots, x_n(X_0,t)) q_i(X_0, t)$$

with initial condition

$$\frac{\mathrm{d}q_i}{\mathrm{d}t}(X_0, t_0) = \begin{cases} 1 & i = 1 \\ 0 & i = 0 \end{cases}$$

First let us look at the initial condition. Note that it doesn't depend on the initial point X_0. Geometrically, consider two points $X_0, X_0 + (\Delta x_{10}, 0, \ldots, 0)$. What we are doing is to take curves passing through these two points at t_0, and looking at the difference of coefficients. As the two curves approach t_0, the first coordinate will depend linearly on the change Δx_{10} while the other coordinates shouldn't change much because the initial points only differ in the first coordinate. This geometric picture is translated into the initial conditions above.

We ask where is the system 2 defined. The system $(q_1, \ldots, q_n)$ have time derivatives of the form $a_1 q_1 + \cdots + a_n q_n$ where a_i are functions of X_0, t. These q_i are themselves functions of X_0, t, however their dependence on X_0 is not as direct as their dependence on t. For a fixed X_0, this is a system defined on $\mathbb{R}^n \times [t_0 \pm a]$, and that is where our solutions lie. The x_i are continuous, differentiable functions $x_i \colon R \to B$ where $B = |X - \tilde{X}_0| \leq b$. The indirect dependence of the q_i on X_0 is captured in the following.

Theorem 9. *Consider a system of differential equations in which the functions f_i depend upon any number of parameters $\mu_1, \ldots, \mu_m$*

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = f_i(x_1, \ldots, x_n; \mu_1, \ldots, \mu_m; t), 1 \leq i \leq n.$$

If each f_i has partial derivatives with respect to $x_1, \ldots, x_n; \mu_1, \mu_2, \ldots, \mu_m$ continuous in some $(n + m + 1)$ -dimensional region R, then the solutions

$$x_i(x_{10}, \ldots, x_{n0}; \mu_1, \ldots, \mu_m; t), 1 \leq i \leq n$$

will have partial derivatives in $\mu_1, \ldots, \mu_m$ continuous in all their arguments through whatever part of R the solutions are defined.

Proof. The idea is to simply treat the parameters as constant functions of t. So, along with the given equations, consider $d\mu_j/dt = 0, 1 \leq j \leq m$. This new system of $n + m$ variables has a solutions with initial value $(x_{10}, \ldots, x_{n0}, \mu_1, \ldots, \mu_m)$ and is continuous with respect to the initial variables. Moreover, by the theorem above, the solutions are differentiable with respect to each coordinate of the initial condition, in particular the μ_js. □

Continuing with the solutions to 2, applying the theorem above, it follows that q_i are continuous as functions of X_0, therefore the solution $(x_1, \dots, x_n)$ is $\mathcal{C}^1$ with respect to the initial value. We can go further. If the f_i are $\mathcal{C}^2$, then because x_i are $\mathcal{C}^1$ with respect to X_0, q_i are differentiable with respect to $x_{10}, \dots, x_{n0}$. From here, it is clear how by increasing the number of variables we can conclude that the solution to $dX/dt = F(X,t)$ is $\mathcal{C}^k$ when F is $\mathcal{C}^k$ with respect to X.

More directly, what we have proved above is that when F is $\mathcal{C}^1$ with respect to the initial conditions, the collection $(x_1, \dots, x_n, q_1, \dots, q_n)$, where x_i, q_j are as above, satisfy a system of differential equations which is $\mathcal{C}^1$ if F is $\mathcal{C}^2$. This is because the x_i have continuous first order partial derivatives. Naturally, a repetition of this process increases the number of variables exponentially, but proves the result for all $\mathcal{C}^k, k < \infty$. The extension to smooth functions is then obvious.

5 Continuous extensions of solutions

This section considers some boundary properties of solutions and can be skipped. Suppose an integral curve is defined in some domain D over the interval $t_0 - l < t_0 < t_0 + m$. We consider what the solution looks like at the end points. Intuitively, they should approach the boundary of D as otherwise we can extend the solution further.

Theorem 10. *Let D be an arbitrary bounded domain (open or closed) in $\mathbb{R}^{n+1}$ and let $F(X,t)$ be continuous in D and satisfy a Lipschitz condition in X locally in D, i.e., aroud each point of D, there is a neighbourhood where the restriction of F is Lipshcitz in X. Suppose X is a solution to $dX/dt = F(X,t)$ passing through (X_0, t_0) defined on the right of t_0 till $t_0 + m$ which cannot be extended further. Then, if $p(t)$ is the distance of the point $(X(t), t)$ from the boundary C of D, then*

$$\lim_{t \to t_0 + m} p(t) = 0.$$

Note that when F is locally Lipschitz in X and D is compact, F is Lipschitz in X.

Proof. Let $\epsilon > 0$ be given and S be the set of those points in D which have a distance $\geq \epsilon$ from C. It is bounded because D is bounded. Suppose as $t \to t_0 + m$, there are infinitely many points in S, so we get an increasing sequence $t_n \to t_0 + m$ such that $P_n = (X(t_n), t_n) \in S$.

Since S is bounded, these P_n have a limit point $\tilde{P} = (\tilde{X}, \tilde{t})$ which is in the interior of D as it can't be in C or in its exterior. Note in particular that $\tilde{t} = t_0 + m$.

Now, there is a rectangle R around $\tilde{P}$ of the form $|X - \tilde{X}| \leq b, |t - \tilde{t}| \leq a$ contained in D. Let M be the bound on $|F|$ in R (finite because F is continuous, R is compact). Furthermore, we can shrink R itself so that F is Lipschitz in X over R.

The idea is to obtain a P_n very close to $\tilde{P}$ in R_1 and then extend the solution to beyond $t_0 + m$. Recall that if we had a rectangle of the form $|X - X_0| \leq b, |t - t_0| \leq a$, then a solution passing through (X_0, t_0) exists on the interval $|t - t_0| \leq h = \min(a, b/M)$.

So, given t_n, to extend it beyond $t_0 + m$, we would like $h > t_0 + m - t_n$. To this end, we first choose P_n such that

$$|\tilde{X} - X(t_n)| \leq b/2, \tilde{t} - t_n \leq a/2.$$

Once we have that, we would have the rectangle

$$|X - X(t_n)| \leq b - |\tilde{X} - X(t_n)|, |t - t_n| \leq a - (\tilde{t} - t_n)$$

contained in R, hence in D and $h = \min(a - (\tilde{t} - t_n), (b - |\tilde{X} - X(t_n)|)/M)$.

Therefore, we initially shrink a so that $a < b/2M$, then $h = a - (\tilde{t} - t_n) > a/2$. Since $\tilde{t} - t_n < a/2$, it is clear that a solution passing through P_n extends beyond $t_0 + m$.

Now we have two solutions passing through P_n defined on a (closed) neighbourhood of t_n, hence they must agree by the uniqueness theorem. Therefore, the original solution can be extended beyond $t_0 + m$ which is contrary to the assumption.

Thus, as $t \to t_0 + m$, the solution must eventually have no points in S. Since $\epsilon > 0$ was arbitrary, the solution must have limit points in the boundary.

Note that since D is bounded, as $t \to t_0 + m$ we always have limit points. □

Remark. What we have shown is that the limit points as $t \to t_0 + m$ (which exist because D is bounded) must all lie in C. It does not mean that there is a single limit point. Moreover, many different solutions may converge to the same limit point in the boundary and this doesn't contradict the uniqueness because the solution is not really defined at the limit point. As an example, restrict the two solutions in the example above on the region $x > 0$ (here f is seen to be locally Lipschitz).

Theorem 11. *If R is the domain $\mathbb{R}^n \times (t_1, t_2)$ and if $F(X,t)$ satisfies a Lipschitz condition uniformly in every subdomain of the type $\mathbb{R}^n \times [t_1', t_2'], t_1 < t_1' < t_2' < t_2$, then the solution of $dX/dt = F(X,t)$ passing through any point of R may be extended throughout the entire open interval (t_1, t_2).*

Proof. Fix a point (X_0, t_0) and look at the maximal interval around it where a solution exists. Let it be (a, b). If $a > t_1$ or $b < t_2$, then one can extend the solution slightly by hypothesis, and patch up the two solutions by uniqueness. Therefore, the solution must exist on (t_1, t_2). Alternately, following [1], one can argue as follows.

Suppose such a solution $X(t)$ cannot be extended beyond some $t_0 < \bar{t} < t_2$. Fix some $\epsilon > 0$ such that $\bar{t} - \epsilon > t_0$, and on $[t_0, \bar{t} - \epsilon]$ consider the approximate solution given by $\tilde{X}(t) \equiv X_0$. The error of approximation is

$$\max_{[t_0, \bar{t}-\epsilon]} |\frac{\mathrm{d}\tilde{X}}{\mathrm{d}t} - F(X_0, t)| \leq \max_{[t_0, \bar{t}]} |F(X_0, t)| = M$$

where a maximum M exists because $F(X_0, t)$ is continuous on the compact set $[t_0, \bar{t}]$.

By the fundamental inequality, we have

$$|X(t) - X_0| \leq \frac{M}{k} e^{k(\bar{t} - t_0) - 1}, t_0 \leq t \leq \bar{t} - \epsilon.$$

Note that the bound is independent of ϵ, so X is bounded to the left of $\bar{t}$, hence by the theorem above, it must approach the boundary of R (for we are considering a maximal solution). However, this is impossible as $\bar{t} < t_2$. Therefore, the solution must exist till t_2. Similarly, the solution must exist to the left of t_0 till t_1. □

6 Vector fields

A note to the reader : Since the following sections are based from a variety of sources there may be certain inconsistencies in the notation. Specifically, the flow along a vector field X may be written as X_t or ϕ_t. However, within each proof, I have tried to be consistent with the notation, so it shouldn't be that much of a problem.

Given a smooth manifold M, the tangent vectors at a point are derivations on the algebra of germs of smooth functions at that point. This collection of derivations forms a vector space (of the same dimension as M) and is called the tangent space. The tangent bundle TM is obtained by taking the disjoint union of all tangent spaces. Using the projection map $\pi \colon TM \to M$, we can topologize TM to arrive at a smooth manifold TM such that π is a smooth map.

Since the tangent space at each point is contractible, TM, M are actually homotopy equivalent. We have the inclusion map $i \colon M \to TM$ with inverse given by π and $i \circ \pi$ is homotopic to the identity map via

$$\begin{aligned} F \colon TM \times I &\to TM \\ ((p, v), t) &\mapsto (p, (1-t)v) \end{aligned}$$

(check smoothness locally). Therefore, from the point of homotopy theory TM, M are identical.

A smooth vector field on an open set U is a section $X: M \to TU$ of the projection map. Given a curve $c \colon I \to M$, its derivative is a vector field on $c(I)$. When talking about smooth curves on closed intervals we always assume that the curve can be extended (smoothly) to some neighbourhood of the closed interval, that way there is no issue of differentiability at the end points.

The question we are interested is the inverse. Given a vector field, can we find curves, passing through some fixed point say, whose derivatives agree with the given vector field. Note here that the derivative of a curve at a point is a linear map from one tangent space to another, since the tangent bundle of the interval is trivialised, we can talk of the derivative as assiging a single tangent vector to each point on the image of the curve.

Locally, after going to a coordinate chart, we are trying to solve a system of differential equations. Let $p \in M$ and $(U, x_1, \ldots, x_n)$ be a chart around p where the vector field is given by some $a_1\partial/\partial x_1 + \cdots + a_n\partial/\partial x_n$, then the curve is going to be of the form $(y_1(t), \ldots, y_n(t))$ such that

$$\frac{\mathrm{d}y_i}{\mathrm{d}t} = a_i(y_i(t)).$$

So, locally it is a question in ordinary euclidean space and we know how to solve it. We are then interested in maximal solutions, curves which can go outside coordinate charts. Any solution will be called an integral curve to the vector field. We have the following theorem from [2].

Theorem 12. *Let X be a $\mathcal{C}^\infty$ vector field on a differentiable manifold M. For each $m \in M$ there exists $a(m), b(m) \in \mathbb{R} \cup \{\pm\infty\}$ and a smooth curve*

$$\gamma_m \colon (a(m), b(m)) \to M$$

such that

(*a*) $0 \in (a(m), b(m))$ *and* $\gamma_m(0) = m$.
(*b*) γ_m *is an integral curve of* X.
(*c*) *If* $\mu \colon (c, d) \to M$ *is a smooth curve satisfying* (*a*), (*b*), *then* $(c, d) \subseteq (a(m), b(m))$ *and* $\mu = \gamma_m|_{(c,d)}$.

We define the following

Definition 3. *For each $t \in \mathbb{R}$ define a transformation X_t with domain*

$$\mathcal{D}_t = \{m \in M : t \in (a(m), b(m))\}$$

by setting $X_t(m) = \gamma_m(t)$.

(*d*) *For each $m \in M$, there exists an open neighbourhood V of m and an $\epsilon > 0$ such that the map*

$$(t, p) \mapsto X_t(p)$$

is defined and is $\mathcal{C}^\infty$ from $(-\epsilon, \epsilon) \times V \to M$.
(*e*) $\mathcal{D}_t$ *is open for each* t.
(*f*) $\cup_{t>0}\mathcal{D}_t = M$.
(*g*) $X_t \colon \mathcal{D}_t \to \mathcal{D}_{-t}$ *is a diffeomorphism with inverse* X_{-t}.
(*h*) *Let s, t be real numbers. Then the domain of $X_s \circ X_t$ is contained in, but not generally equal to, $\mathcal{D}_{s+t}$. However, the domain of $X_s \circ X_t$ is $\mathcal{D}_{s+t}$ in the case in which s, t both have the same sign. Moreover, on the domain of $X_s \circ X_t$, we have $X_s \circ X_t = X_{s+t}$.*

Before going for the proof, let us discuss what the theorem is saying. We have a vector field X. From earlier considerations we know that we can find local solutions. Moreover, by smooth dependence on initial conditions we know that locally the curves depend smoothly on the initial parameter. Now we can consider the maximal possible solution (because we know that solutions exist). The operation X_t sends $\mathcal{D}_t$ forward (or backward) in time along X. And statement h says that going ahead $s + t$ units is the same as going t units and then s units of time. As an example, for a particular vector field on the sphere, we would get a slowly rotating ball, X_t would be the family of rotations parametrised by time.

Proof. We take $(a(m), b(m))$ to be the maximal interval around 0 such that a solution exists passing through m at $t = 0$. It is non empty by the existence theorem (and this proves (f)). Moreover, by the uniqueness theorem (c) follows.

Fix a point m and let $(U, x_1, \ldots, x_n)$ be a coordinate chart around m. We translate everything to U, so we have smooth functions $a_1, \ldots, a_n$ defined on U which we can treat as being defined on $U \times \mathbb{R}$.

Now, by using previous theorems and shrinking U to a rectangle if necessary, we know that there is some $\epsilon > 0$ and a neighbourhood V of m where a solution exists on $V \times (-\epsilon, \epsilon)$ and is smoothly dependent on the initial point. Lifting everything back to M proves (d).

Let $t \in (a(m), b(m))$, then $s \mapsto \gamma_m(t+s)$ is an integral curve of X on the maximal domain $(a(m) - t, b(m) - t)$ with initial condition $0 \mapsto \gamma_m(t)$. The interval is maximal because otherwise we could extend $(a(m), b(m))$. It then follows, by (c), that

$$(a(m) - t, b(m) - t) = (a(\gamma_m(t)), b(\gamma_m(t)))$$

and for s in the above interval,

$$\gamma_{\gamma_m(t)}(s) = \gamma_m(t+s).$$

Now, if m is in the domain of $X_s \circ X_t$, then $t \in (a(m), b(m))$ and from the above, $s \in (a(m)-t, b(m)-t)$, i.e., $s + t \in (a(m), b(m))$. Thus, $m \in \mathcal{D}_{s+t}$ and we have $X_s(X_t(m)) = X_{s+t}(m)$. When s, t have the same sign and $m \in \mathcal{D}_{s+t}$, then we immediately have $t \in (a(m), b(m))$, hence $s \in (a(\gamma_m(t)), b(\gamma_m(t)))$, hence m is in the domain of $X_s \circ X_t$.

Parts (e), (g) are trivial for $t = 0$, so we assume $t > 0$ (the negative case is similar). Let $m \in \mathcal{D}_t$. From (d), for every $p \in \gamma_m([0, t])$, there is a neighbourhood V_p of p and an $\epsilon_p > 0$ such that $X_s(q)$ is defined on $(-\epsilon_p, \epsilon_p) \times V_p$. Cover $\gamma_m([0, t])$ by these V_p and obtain a finite subcover whose union is W and take the minimum of the ϵs associated to the finite subcover to get an $\epsilon > 0$ such that $X_s(q)$ is defined on $(-\epsilon, \epsilon) \times W$.

What this means is that we can smoothly move ϵ units of time for every point in W. Now, we need to obtain a neighbourhood around m where we can move ahead t units of time. To this end, obtain an n large enough so that $t/n < \epsilon$. The idea is that we move ahead by t/n units and if we are still in W, then we can move ahead some more.

Note that $X_{t/n}$ is defined on W and is a smooth function to M. Now, set $\alpha_1 = X_{t/n}|_W$ and $W_1 = \alpha_1^{-1}(W)$. Then for $i = 2, \ldots, n$ we inductively define

$$\alpha_i = X_{t/n}|_{W_i} \text{ and } W_i = \alpha_i^{-1}(W_{i-1})$$

(alternately one can just take the inverse under $X_{t/n}$ and intersect the inverse with W_{i-1}). Each α_i is $\mathcal{C}^\infty$ and hence W_n is an open subset of W containing m (it contains m as $\gamma_m(t) \in W$), and by part (h),

$$\alpha_1 \circ \cdots \circ \alpha_n|_{W_n} = X_t|_{W_n}.$$

Consequently, $W_n \subseteq \mathcal{D}_t$, hence $\mathcal{D}_t$ is open proving (e).

Note that above we have obtained X_t as a composition of smooth maps locally, hence X_t is a smooth map for all $t \in \mathbb{R}$. By part (h), X_t, X_{-t} are inverses of each other on their respective domains. Therefore, X_t maps $\mathcal{D}_t$ diffeomorphically onto $\mathcal{D}_{-t}$ proving part (g). □

We do not always have the domain of $X_s \circ X_t$ as $\mathcal{D}_{s+t}$, consider for example $M = (0, \infty)$ and the vector field d/dt, then from each point we can move indefinitely to the right, but only for a finite time to the left, therefore the domain of composition of $X_t \circ X_{-t}$ is not the whole space for $t > 0$, but $X_0 = (0, \infty)$.

Definition 4. *A smooth vector field X on M is complete if $\mathcal{D}_t = M$ for all t. In this case, the transformations X_t form a group of transformations of M parametrized by the real numbers called the 1-parameter group of X.*

Theorem 13. *Smooth vector fields on compact manifolds are complete.*

Proof. We repeat an argument from the previous theorem. Let M be a compact manifold and X a smooth vector field, around every point $m \in M$ there is a neighbourhood W_m and an $\epsilon_m > 0$ such that a solution to X exists on $(-\epsilon_m, \epsilon_m) \times W_m$. Cover M by W_m and obtain a finite subcover, then it follows that there is an $\epsilon > 0$ such that a solution exists on $(-\epsilon, \epsilon) \times M$. Given $t > 0$ (the negative case is similar), we take an n large enough so that $t/n < \epsilon$. Now each point can be moved ahead t/n units, composing n times we see that $\mathcal{D}_t = M$, i.e., X is complete. □

Given a smooth vector field X on a manifold M, we say that a point $p \in M$ is regular if $X_p \neq 0$ and singular otherwise.

Theorem 14. (*Canonical form for a Regular Vector Field*) *Let X be a smooth vector field on a manifold M and let $p \in M$ be a regular point. There exist coordinates $(x_1, \ldots, x_n)$ on a neighbourhood of p where X is equal to $\partial/\partial x_1$.*

Proof. What this means is that the diffeomorphism provided by the chart makes the x_1-lines into the flows of X. Let $(U, y_1, \ldots, y_n)$ be a chart near p with p corresponding to origin and by a suitable linear transformation we may assume that $X_p = \partial/\partial y_1$. There is a neighbourhood U_0 of p and an $\epsilon > 0$ such that a flow Φ of X is defined on $(-\epsilon, \epsilon) \times U_0$ and its image lies in U. By shrinking U_0 we may assume that it is a rectangle of the form $I \times S$ where I is an interval and S is an $n-1$ dimensional rectangle.

Define the smooth map

$$\begin{aligned} \psi \colon (-\epsilon, \epsilon) \times S &\to U \\ (t, q) &\mapsto \Phi(t, (0, q)) \end{aligned}$$

We are essentially using the "zero section" of the projection $U_0 \to S$ which is why we needs the assumption that U_0 is rectangular. Now,

$$\begin{aligned} (\psi_* \frac{\partial}{\partial t}|_{(t_0, x_0)}) f &= \frac{\partial}{\partial t}|_{(t_0, x_0)} (f \circ \psi) \\ &= \frac{\partial}{\partial t}|_{t = t_0} f(\Phi(t, (0, x_0))) \\ &= X_{\psi(t_0, x_0)} f \end{aligned}$$

because Φ is the flow corresponding to X. At the same time we have

$$\psi(0, x_2, \ldots, x_n) = \Phi(0, (0, x_1, \ldots, x_n)) = (0, x_2, \ldots, x_n)$$

therefore,

$$\psi_* \frac{\partial}{\partial x_i}|_{(0,0)} = \frac{\partial}{\partial y_i}|_{(0,0)}$$

hence at the origin, ψ_* maps a basis to a basis, hence applying the inverse function theorem we obtain a diffeomorphism $\psi \colon W \to Y$ where W is a neighbourhood of $(0,0)$ and Y is a neighbourhood of p. The inverse is a chart where the pullback of $\partial/\partial t$ is X thus completing the proof.

Geometrically, we take the horizontal slice of U_0 and send all the points in this slice "upwards" along X (imagining x_1 to be upwards). Then, a neighbourhood of p is parameterised by it's position in the horizontal slice and a time coordinate along the flows of X. Intuitively, ψ takes the horizontal slice and translates it along X (which we are imaginging as not lying on the horizontal slice). □

The interesting things happen around singular points. When p is a regular point, the flow behaves like a steady stream moving in the x_1 direction, but when it is singular, while p is fixed, points around it may swirl around p, asymptotically approach p or go off to infinity and so on. This sort of behaviour is studied under the name of smooth dynamical systems.

It is very tempting to want a similar canonical form for a collection of vector fields $X_1, \ldots, X_m$ near p. Above, we fixed one slice and translated it in time. Ideally, the general case would involve fixing some $n-m$ dimensional slice and translating that along m flows (one for each vector field) for varying amounts of time to reach a different point. This way we could parametrize a neighbourhood using m time coordinates and $n-m$ space coordinates. However, a problem occurs when we look at the order of the flows we use to translate our slice and for it to be order independent we will need the vector fields to be "involutive".

Now, let X be a smooth vector field on a manifold M and fix a point $p \in M$. There is a neighbourhood $(-\epsilon, \epsilon) \times V$ of $(0, p)$ and a map $\Phi \colon (-\epsilon, \epsilon) \times V \to M$ which defines the flow corresponding to

X, i.e., for each $q \in V$, the curve $t \mapsto \Phi(t,q), t \in (-\epsilon, \epsilon)$ is an integral solution to X passing through q. We study the derivative of Φ.

For a fixed t, denote by j_t the inclusion $V \hookrightarrow \{t\} \times V$. Then, $X_t = \Phi \circ j_t, t \in (-\epsilon, \epsilon)$. By shrinking V if necessary we may assume that V has coordinates and that Φ lands in an open set with the same coordinates (for example, obtain V and consider $\Phi^{-1}(V)$). Then in local coordinates (we don't need explicit coordinates, but it helps to think in terms of coordinates), the derivative of Φ is of the form

$$\begin{bmatrix} X(\Phi(t,q)) & DX_t(q) \end{bmatrix}_{n+1 \times n}$$

Fix a time t and suppose $t \pm \delta \subset (-\epsilon, \epsilon)$. Consider the map

$$\begin{aligned} F \colon (-\delta, \delta) \times W &\to M \\ (t', q) &\mapsto \Phi(t + t', q) \end{aligned}$$

where W is some smaller neighbourhood so that F lands in V. This is a well defined, smooth map and equals $\Phi \circ j_t \circ \Phi$ on $(-\delta, \delta) \times V$ because $X_t \circ X_{t'} = X_{t+t'}$ (we need to shrink V so that the composition is defined).

The image of the vector $\partial/\partial t$ at $(0,p)$ under F is going to be $X(\Phi(t,p))$, while the derivative taken along the composition $\Phi j_t \Phi$ is going to be, by chain rule, $DX_t(p)(X(p))$. Therefore, we have

$$(X_t)_*(p)(X(p)) = X(\Phi(t,p))$$

in other words, the push forward of $X(p)$ along the flow X_t is going to be $X(X_t(p))$. Since t, p were arbitrary, the result holds at all other times and points also. Next we look at the push forward of other vectors. In order to do so, we introduce Lie Brackets.

6.1 Lie Bracket

On $\mathbb{R}^n$ we know that the partial derivatives commute on smooth functions. On manifolds, it doesn't make sense to compose two derivatives at a point because derivatives act on functions. Therefore, we must consider vector fields acting on functions to obtain other functions, and in fact this is what happens over $\mathbb{R}^n$ as well, where we consider constant vector fields.

Given two vector fields X, Y on a manifold, and a smooth function $f \colon M \to \mathbb{R}$, we have a smooth function Xf. We can then act Y on Xf to get $Y(Xf)$. In this way we can compose two vector fields, however YX is not itself a vector field. We look at the commutator

$$[X, Y] = XY - YX$$

called the Lie Bracket of X, Y.

It is easy to verify that $[X, Y]$ is another vector field and we have (see [2])

(a) $[X, Y]$ is a smooth vector field on M.
(b) If $f, g \in \mathcal{C}^\infty(M)$, then $[fX, gY] = fg[X,Y] + f(Xg)Y - g(Yf)X$.
(c) $[X, Y] = -[Y, X]$.
(d) (Jacobi identity) $[[X,Y],Z] + [[Y,Z],X] + [[Z,X],Y] = 0$ for all smooth vector fields X, Y, and Z on M.

Suppose we have a coordinate chart $(U, x_1, \ldots, x_n)$ and two vector fields given by

$$X = \sum a_i \frac{\partial}{\partial x_i}, Y = \sum b_i \frac{\partial}{\partial x_i}.$$

Now given a germ f we compute the derivative at a point p as follows

$$\begin{aligned} [X,Y]_p(f) &= X_p(Yf) - Y_p(Xf) \\ &= \sum_i a_i(p) \frac{\partial}{\partial x_i}\bigg|_p \left(\sum_j b_j \frac{\partial f}{\partial x_j} \right) - \sum_i b_i(p) \frac{\partial}{\partial x_i}\bigg|_p \left(\sum_j a_j \frac{\partial f}{\partial x_j} \right) \\ &= \sum_{i,j} \left(a_i(p) \frac{\partial b_j}{\partial x_i}(p) - b_i(p) \frac{\partial a_j}{\partial x_i}(p) \right) \frac{\partial f}{\partial x_j} \end{aligned}$$

where we have used the fact that for smooth functions (or more generally $\mathcal{C}^2$ functions) the double derivatives commute (i.e., the lie bracket of the constant vectors in regular euclidean space is zero). Note in particular what happens if Y is a constant vector for example, i.e., if the b_is are constant.

Definition 5. *Let $\varphi\colon M \to N$ be $\mathcal{C}^\infty$. Smooth vector fields X on M and Y on N are called φ-related if $d\varphi \circ X = Y \circ \varphi$ (wherever it makes sense, i.e., in the image of φ). In other words, Y is the push forward of X.*

Theorem 15. *Let $\varphi\colon M \to N$ be $\mathcal{C}^\infty$. Let X and X_1 be smooth vector fields on M and let Y and Y_1 be smooth vector fields on N. If X is φ-related to Y and X_1 is φ-related to Y_1, then $[X, X_1]$ is φ-related to $[Y, Y_1]$, i.e., Lie bracket commutes with push forwards..*

Proof. Let $m \in M$ and f a $\mathcal{C}^\infty$ germ at $\varphi(m)$. We need to show that

$$d\varphi([X, X_1]_m)(f) = [Y, Y_1]_{\varphi(m)}(f).$$

Just applying the definitions, we have

$$\begin{aligned}
d\varphi([X, X_1]_m)(f) &= [X, X_1]_m(f \circ \varphi) \\
&= X_m(X_1(f \circ \varphi)) - X_1|_m(X(f \circ \varphi)) \\
&= X_m((d\varphi \circ X_1)(f)) - X_1|_m((d\varphi \circ X)(f)) \\
&= X_m(Y_1(f) \circ \varphi) - X_1|_m(Y(f) \circ \varphi) \\
&= d\varphi(X_m)(Y_1(f)) - d\varphi(X_1|_m)(Y(f)) \\
&= Y_{\varphi(m)}(Y_1(f)) - Y_1|_{\varphi(m)}(Y(f)) \\
&= [Y, Y_1]_{\varphi(m)}(f)
\end{aligned}$$

□

Now, let X, Y be smooth vector fields on M and fix a point $p \in M$. Let V be a neighbourhood of p such that a flow Φ is defined on $(-\epsilon, \epsilon) \times V$ for the vector field X.

Theorem 16. *If $[X, Y] = 0$, then $\Phi_* Y = Y$, i.e., for all t, $X_{t*}(Y) = Y$.*

Proof. We're seeing Y as a vector field on $(-\epsilon, \epsilon) \times V$ by setting the time coefficient to 0, i.e., the unique vector field related to Y via projection. We know that at time $t = 0$ $\Phi_* Y = Y$ because the flow is the identity map. The idea is to compute the time derivative of $\Phi_* Y$ and show that it is zero. If we think in terms of coordinates, they are a priori functions of time and position and we hope to show that they infact only depend on the position, i.e, $\Phi(t, q)$. We introduce a time vector to distinguish the same point at different times, so that we can see the pushforward of Y at a single point at different times: extend Φ to $\tilde{\Phi}$ by

$$\begin{aligned}
\tilde{\Phi}\colon (-\epsilon, \epsilon) \times V &\to (-\epsilon, \epsilon) \times M \\
(t, q) &\mapsto (t, \Phi(t, q))
\end{aligned}$$

We compute the bracket $[\partial/\partial t, \tilde{\Phi}_* Y]$. Because $\partial/\partial t$ is a constant vector field, this bracket gives us the time derivative of Y. Note that

$$(\tilde{\Phi}_* \frac{\partial}{\partial t})(t, \Phi(t, q)) = \frac{\partial}{\partial t}(t, \Phi(t, q)) + X(\Phi(t, q))$$

by using coordinates (look at the image of the t-curves in $(-\epsilon, \epsilon) \times V$). Using $\tilde{\Phi}_* X = X$, we have

$$\begin{aligned}
[\frac{\partial}{\partial t}, \tilde{\Phi}_* Y]_{(t, \Phi(t,q))} &= [\tilde{\Phi}_* \frac{\partial}{\partial t} - \tilde{\Phi}_* X, \tilde{\Phi}_* Y]_{(t, \Phi(t,q))} \\
&= \tilde{\Phi}_* [\frac{\partial}{\partial t}, Y]_{(t,q)} - \tilde{\Phi}_* [X, Y]_{(t,q)}
\end{aligned}$$

at every point $(t, \Phi(t, q)) \in (-\epsilon, \epsilon) \times M$. Since Y is independent of time, the first term is zero and the second is zero by assumption. So, the bracket $[\partial/\partial t, \tilde{\Phi}_* Y] = 0$ wherever it is defined.

Again, on $(-\epsilon, \epsilon) \times M, \partial/\partial t$ is a constant vector field, so the left hand side is just the vector field obtained by time differentiating the coefficients of $\tilde{\Phi}_* Y$. $\tilde{\Phi}_* Y$ is a function of (t, q_1) and the above says that it is independent of t, meaning that the pushforward of Y at some point q_1 doesn't depend on where you arrived from (i.e., the starting point and time taken to reach q_1 along X).

Since the time derivative is zero, the coefficients of $\tilde{\Phi}_* Y$ at any (t, q_1) in the image of $\tilde{\Phi}$ is independent of t and equals the value at $(0, q_1)$ which is $Y(q_1)$. It then follows that for every time t, $X_{t*}Y = Y$ as required. □

For another proof, see [3]. To recap, the idea is to show that the pushforward of Y is independent of time, hence the curve so described in $T_{q_1}M$ is constant. Since at $t = 0, X_t$ is identity, this means that the pushforward is Y itself.

Geometrically, fix a point q_1 and imagine transporting the vector field Y along X from various starting points. This gives a collection of vectors at q_1, varying in time (intuitively, it is a function of time because the starting point can be recovered by going back from q_1 for t time), in other words, a path in the tangent space at q_1. The above says that this is the constant path. More generally, we have shown here that

$$[\frac{\partial}{\partial t}, \tilde{\Phi}_* Y]_{(t,\Phi(t,q))} = -\tilde{\Phi}_*[X, Y]_{(t,q)} \tag{3}$$

This formula gives the velocity (not quite) of the path mentioned above, i.e., the Lie bracket (or rather, its pushforward along the flow to the point in question) gives the derivative of Y along X. So, to get the pushforward of Y, we need to integrate the Lie bracket along a path (technically we need to pushforward and then integrate, so that all vectors are in the same tangent space).

Next we show that the flows of commuting vector fields themselves commute. If we have two commuting vector fields X, Y with flows ϕ_t, ψ_t respectively we show that $\phi_t \circ \psi_s = \psi_s \circ \phi_t$ wherever both sides are defined. Geometrically this encloses a parallelogram, going t along X and s along Y is the same as going along Y first and then X. Before that we prove the following more general result, from [3].

Theorem 17. *Suppose $F\colon M \to N$ is a smooth map, X, Y smooth vector fields on M, N respectively with corresponding flows ϕ, ψ. Then X and Y are F-related if and only if for each $t \in \mathbb{R}, \psi_t \circ F = F \circ \phi_t$ on the domain of ϕ_t.*

Proof. Denote by $\mathcal{D}_{t,X}$ the domain of ϕ_t and similarly $\mathcal{D}_{t,Y}$. First, suppose X, Y were F-related, i.e., $F_{*,p}X_p = Y_{F(p)}\ \forall p$. Let γ be a maximal flow through p, then

$$\begin{aligned}(F \circ \gamma)'(t) &= F_{*,\gamma(t)}(\gamma'(t)) \\ &= F_{*,\gamma(t)}X_{\gamma(t)} \\ &= Y_{F(\gamma(t))}\end{aligned}$$

So, $F \circ \gamma$ is a flow through $F(p)$ and by uniqueness results, we see that the flow through $F(p)$ is defined for at least as much time as it is defined for p. It then follows that $\psi_t \circ F = F \circ \phi_t$ on $\mathcal{D}_{t,X}$.

Conversely, suppose $F(\mathcal{D}_{t,X})$ lands in $\mathcal{D}_{t,Y}$ and that $\psi_t \circ F = F \circ \phi_t$ for all t. Fix a point p and let γ be a (maximal) flow through p defined on a domain I_p. Then for all $t \in I_p, \gamma(t) = \phi_t(p)$, so $F \circ \gamma(t) = \psi_t \circ F(p)$ is the flow through $F(p)$ by definition. So,

$$\begin{aligned}F_{*,p}X_p &= F_{*,p}(\gamma'(0)) \\ &= (F \circ \gamma)'(0) \\ &= Y_F(p)\end{aligned}$$

completing the proof. □

Theorem 18. *Let M be a smooth manifold, X, Y smooth vector fields on M with flows φ, ψ defined on some neighbourhood of $(0, p), p \in M$. Then X, Y commute if and only if their flows commute, i.e., $\varphi_t \circ \psi_s = \psi_s \circ \varphi_t$ wherever defined.*

Proof. Suppose that the flows of X, Y are defined on a neighbourhood $(-\epsilon, \epsilon) \times V$ of $(0, p)$. If X, Y commute, i.e., $[X, Y] = 0$, then we know that $\varphi_{t*}Y = Y$, i.e., Y is invariant under φ_t for all t, which means that Y is φ_t-related to Y. By the theorem above,

$$\varphi_t \circ \psi_s = \psi_s \circ \varphi_t$$

where we use the same notation ψ_s for flows on V, M because they are the flows of the same vector field.

Conversely, if the flows commute, Y is φ_t-related to Y on M, which means that the time derivative of $\varphi_{t*}Y$ at a fixed point is zero, so $[X, Y] = 0$ by Equation 3. □

With this result, we can prove a canonical form for commuting vector fields as we shall see next.

6.2 Distributions and Frobenius Theorem

Definition 6. *Let c be an integer, $1 \le c \le d$. A c-dimensional distribution $\mathscr{D}$ on a d-dimensional manifold M s a choice of a c-dimensional subspace $\mathscr{D}(m)$ of M_m for each $m \in M$. $\mathscr{D}$ is smooth if for each $m \in M$ there is a neighbourhood U of m and there are c vector fields $X_1, \ldots, X_c$ of class $\mathcal{C}^\infty$ on U which span $\mathscr{D}$ at each point of U. A vector field X on M is said to belong to (or lie in) the distribution $\mathscr{D}(X \in \mathscr{D})$ if $X_m \in \mathscr{D}(m)$ for each $m \in M$. A smooth distribution $\mathscr{D}$ is called involutive (or completely integrable) if $[X, Y] \in \mathscr{D}$ whenever X, Y are smooth vector fields lying in $\mathscr{D}$.*

Definition 7. *A submanifold (N, ψ) of M is an integral manifold of a distribution $\mathscr{D}$ on M if $d\psi(N_n) = \mathscr{D}(\psi(n))$ for each $n \in N$.*

Here a submanifold refers to a manifold N together with an **injective immersion** $\psi \colon N \to M$. It is clear that finding an integrable manifold for a given distribution is a generalization of finding integral curves satisfying a given vector field, the only difference being that integral curves are supposed to match the vector field, whereas integrable manifolds need only span the same subspace as the distribution.

Remark. (Non rigorous, intuitive remark) In the one dimensional case, given a vector field, consider the induced distribution. Now, the derivative of local parametrizations of integral submanifold is going to be scalar multiples of the given vector field and this is going to be a smooth function on the parameter space. Using a small change of variable, the parametrization can be made into an integral curve of the vector field.

Lemma 6. *Let $\mathscr{D}$ be a smooth disctribution on M such that through each point of M there passes an integral manifold of $\mathscr{D}$. Then $\mathscr{D}$ is involutive.*

Proof. Let X, Y be smooth vector fields in $\mathscr{D}$ and let $m \in M$. Let (N, ψ) be an integral manifold of $\mathscr{D}$ through m, and suppose $\psi(n_0) = m$. Since $d\psi \colon N_n \to \mathscr{D}(\psi(n))$ is an isomorphism (because ψ is an immersion) at each $n \in N$, there exists vector fields $\tilde{X}, \tilde{Y}$ on N such that $d\psi \circ \tilde{X} = X \circ \psi$, $d\psi \circ \tilde{Y} = Y \circ \psi$. Then, the pushforward of the bracket $[\tilde{X}, \tilde{Y}]$ is the bracket $[X, Y]$, hence $[X, Y]_m \in \mathscr{D}$ and $\mathscr{D}$ is involutive. □

Theorem 19. (*Canonical form for commuting vector fields*) *Let M be a smooth n-manifold and $X_1, \ldots, X_m$ be a collection of smooth commuting vector fields defined on an open subset U such that at each $p \in U$, the vectors are linearly independent. Then there exists a coordinate chart $(V, x_1, \ldots, x_n)$ around p such that $X_i = \partial/\partial x_i, 1 \le i \le m$ on V.*

Proof. First find a chart $(V, y_1, \ldots, y_n$ around $p \in U$ such that p maps to the origin. Let $\phi_1, \ldots, \phi_m$ denote the flows of $X_1, \ldots, X_m$ respectively. By shrinking various neighbourhoods, arrive at a neighbourhood U_0 of p where $(\phi_m)_{t_m} \circ \cdots \circ (\phi_1)_{t_1}$ are defined on U_0 when all $|t_i| < \epsilon$ for some $\epsilon > 0$ and the image lands in V. We can shrink U_0 further and assume that it is a rectangle of the form $I \times S$ where I is m-dimensional and $S, n - m$ dimensional, centred at the origin.

Consider the map

$$\begin{aligned} F\colon (-\epsilon,\epsilon)^m \times S &\to U \\ (t_1,\dots,t_m,q_{m+1},\dots,q_n) &\mapsto (\phi_m)_{t_m}\circ\cdots\circ(\phi_1)_{t_1}(0,\dots,0,q_{m+1},\dots,q_n) \end{aligned}$$

where we are again using the section $S \to U_0$. This is clearly a smooth map.

Since the distribution is involutive, it is easy to see that the pushforward of $\partial/\partial t_i$ is X_i everywhere, because

$$(\phi_m)_{t_m}\circ\cdots\circ(\phi_1)_{t_1}(0,q) = (\phi_i)_{t_i}((\phi_m)_{t_m}\circ\cdots\circ\widehat{(\phi_i)_{t_i}}\circ\cdots\circ(\phi_1)_{t_1}(0,q))$$

so we get an integral curve through some point for X_i meaning that the pushforward of $\partial/\partial t_i$ from (t,q) is indeed X_i at $F(t,q)$ where $t=(t_1,\dots,t_m), q=(q_{m+1},\dots,q_n)$.

Next, at the origin, $F(0,q)=q$ so the pushforward of the $\partial/\partial q_i$ vectors is just $\partial/\partial q_i$ which means that at $0, F$ is invertible. By the inverse function theorem, we obtain a chart with the required properties. ☐

As a corollary, if we have a distribution which is spanned by commuting vector fields (at least locally), then locally we can find a chart where the slices, obtained by setting the last $n-m$ coordinates constant, are integrable submanifolds. In [3] such charts are called "flat" charts, other use "adapted" charts. The following theorem of Frobenius is one of the most important results in the theory of smooth manifolds.

Theorem 20. (*Frobenius*) *Given an m-dimensional smooth distribution $\mathscr{D}$ on an n-dimensional manifolds M, around each point p there is a chart diffeomorphic to a cubical neighbourhood around the origin such that the slices are integral manifolds of $\mathscr{D}$. Moreover, if (N,ψ) is a connected integral manifold contained in U, then it must lie in one of these slices.*

Proof. The first part amounts to finding a basis of $\mathscr{D}$ consisting of commuting vector fields. The following argument is borrowed from [5].

Pick a chart $(U,y_1,\dots,y_n)$ around p where a smooth frame $Y_1,\dots,Y_m$ for $\mathscr{D}$ exists, and after a linear change of variables we may assume that $Y_i|_p = \partial/\partial y_i|_p$. In this neighbourhood write $Y_i=\sum_j a_{ij}\partial/\partial y_j$ and set $A=(a_{ij})_{1\le i,j\le r}$. A is a smoothly varying matrix on U and is non singular at p, hence on a neighbourhood V of p (V can be smaller than U because V only captures those points where the first $r\times r$ minor is non singular). Let B denote the smoothly varying inverse and consider the new frame $X_i = BY_i$.

Since $\mathscr{D}$ is a distribution, $X_i\in\mathscr{D}$ and moreover, it is a frame on V because B is invertible. On V,

$$X_i = \frac{\partial}{\partial x_i} + \sum_{k=r+1}^{n} c_{ik}\frac{\partial}{\partial x_k}, 1\le i\le r$$

for some smooth functions c_{ik}.

If we were to compute the Lie brackets we get

$$[X_i,X_j] = [\frac{\partial}{\partial x_i},\frac{\partial}{\partial x_j}] + \sum_{k=r+1}^{n}[\frac{\partial}{\partial x_i}, c_{jk}\frac{\partial}{\partial x_k}] + \sum_{k=r+1}^{n}[c_{ik}\frac{\partial}{\partial x_k},\frac{\partial}{\partial x_j}] + \sum_{k,k'=r+1}^{n}[c_{ik}\frac{\partial}{\partial x_k}, c_{jk'}\frac{\partial}{\partial x_{k'}}].$$

Observe that this is in the span of $\partial/\partial x_k, k\ge r+1$. However, since $\mathscr{D}$ is involutive, it must be a linear combination of $X_1,\dots,X_r$. But this is possible only if all the brackets are 0, therefore, $X_1,\dots,X_r$ commute.

So, we have found a local frame of commuting vector fields and by using the canonical form for commuting vector fields, we have the first part of Frobenius' theorem. Note that once we have set a chart, all other computations are essentially happening in Euclidean space.

Lastly, suppose $\psi\colon N\to U$ is an injective immersion which is an integral manifold for $\mathscr{D}$ with N connected. If π denotes projection onto the last $n-m$ coordinates, then $d\pi$ annihilates $\mathscr{D}$, hence $\psi\circ\pi$ is a constant map on N (because N is connected, else it is constant on the path components of N), which means that N has to land in one of the slices, S. Now, $N\to S$ is an injective immersion between manifolds of the same dimension, hence a local diffeomorphism, so an open map and a homeomorphism into an open subset of S.

If N were not connected, then it is a countable union (by second countability axiom) of its connected components, hence it's image must lie in a countable union of slices. $\square$

Remark. [2] has a different proof of the Frobenius theorem using induction where the basic idea is to use the canonical form for one vector, apply it to Y_1 and then remove the Y_1 component from the other $m-1$ vectors. Restrict these to the slice $y_1 = 0$ (we start with a rectangular centred chart) and show that the distribution so obtained is involutive (it works out because Lie brackets and pushforwards commute - pushforward with respect to inclusion of the slice in the chart). Using the inductive hypothesis, obtain a rectangular $m-1$-dimensional chart around the origin and then add the y_1 coordinate onto it (it requires some work to show that one can simply add the y_1 coordinate).

6.3 Maximal integral manifolds

Definition 8. *A maximal integral manifold (N, ψ) of a distribution $\mathscr{D}$ on a manifold M is a connected integral manifold of $\mathscr{D}$ whose image is not a proper subset of any other connected integral manifold of $\mathscr{D}$.*

This is just an extension of the notion of a maximal integral curve. In the one dimensional case, we had the notion of a maximal interval containing the origin using which we could patch up the solutions. Here the situation is slightly different. Note that in the one dimensional case we had curves passing through p, but here we just have regions containing p, and we want to patch up the differentiable structures on the integral manifolds.

Lemma 7. *Suppose $\psi\colon N \to M$ is a smooth map factoring through a $\phi\colon P \to M$ which is an integral manifold for a distribution $\mathscr{D}$, i.e., $\psi(N) \subseteq \phi(P)$. Let $\psi_0\colon N \to P$ be the unique map such that $\phi \circ \psi_0 = \psi$, then ψ_0 is smooth.*

Proof. Fix an open $U \subset P$ and pick a point $p \in U$. Let $(V, x_1, \dots, x_n)$ be a cubic centred neighbourhood around $\phi(p)$ whose slices are integral manifolds of $\mathscr{D}$ and let $\tilde{U}$ be a neighbourhood of p landing into V such that $\phi(\tilde{U})$ is the slice $x_{m+1} = \dots = c_n = 0$ (just by taking a preimage). Let $\psi_0(p_1) = p$.

Suppose we prove continuity, then by inverse function theorem, a projection π of V gives a chart $\pi \circ \phi$. Continuity gives as an open neighbourhood around p_1 such that $\pi \circ \phi \circ \phi_0$ is smooth on that neighbourhood, giving us smoothness (for more detail, consult [2], Theorem 1.32).

Let W be the component of $\psi^{-1}(V)$ containing p_1. To show continuity, we just need to prove that $\psi_0(W) \subseteq \tilde{U}$. Because ϕ is injective, it suffices to show that $\psi(W)$ is in a slice of V. Observe that $\psi(W)$ is connected and intersects the slice above.

Moreover $\psi(W)$ lies in a component of $\phi(P) \cap V$, and we have seen that the components of $\phi(P) \cap V$ are contained in slices of V, and this completes the proof. $\square$

Theorem 21. *Let $\mathscr{D}$ be an m-dimensional distribution on an n-dimensional manifold M. Through each $p \in M$ there passes a unique maximal connected integral manifold of $\mathscr{D}$, and every connected integral manifold of $\mathscr{D}$ through p is contained in the maximal one.*

The following proof is completely borrowed from [2].

Proof. First we prove existence. Given p, let K be the set of all points x for which there is a piecewise smooth path from p to x whose smooth parts are integral curves of $\mathscr{D}$ (i.e., tangents lie in the distribution). Note that K is connected and in a cubical neighbourhood, as in the canonical form, around $x \in K$, the appropriate slice containing x is contained in K.

By second countability of M, cover K by a countable collection $\{(U_i, x_1^1, \dots, x_n^i)\}$ of cubical neighbourhoods whose slices

$$x_{m+j}^i = \text{constant}, 1 \leq j \leq n - m$$

are integral manifolds for $\mathscr{D}$. Each slice is Euclidean with coordinates $x_1, \dots, x_m$.

We prove that K, with a suitable topology is a maximal integral manifold through p. First the topology, the collection of all slices which are integral manifolds for $\mathscr{D}$ will be taken as coordinate charts for K and call a subset V open if its intersection with each such slice is open in the slice (i.e., we want the inclusion of each slice to be a continuous map).

Each slice is homeomorphic to it's image in K because in M the intersection of the slices carries a "coherent" topology. This also proves that K is Hausdorff because either two slices are disjoint, or their intersection is a Euclidean space, hence Hausdorff (the slices are open in K).

Ignoring second countability for the moment, we now have a manifold K with an inclusion $K \hookrightarrow M$ which is an injective immersion and an integral manifold for $\mathscr{D}$ because locally, we can go to the slices, we have integral manifolds. Once we prove second countability, we have a path connected (by construction) integral manifold.

To prove second countability, observe that K is covered by countably many charts as above, hence it suffices to prove that K intersects each cubic neighbourhood as above in countably many slices. Let p be contained in $(U_0, x_1^0, \ldots, x_n^0)$ and fix another such chart $(U_i, x_1^i, \ldots, x_n^i)$. We want to show that $K \cap U_i$ is countable union of slices of U_i. Each point $x \in U_i$ is reachable by a piecewise smooth curve from p lying in K and by its compactness we can cover such a curve by a (not necessarily unique) finite sequence of charts $U_0, U_{i_1}, \ldots, U_{i_r}, U_i$.

There are countably many such sequences. Now observe that if S is one slice and U_k is one of these cubic charts, then $S \cap U_k$ is an open subset of second countable S, hence has finitely many components and each component is a connected integral manifold for $\mathscr{D}$, which means that each component has to be in a slice of U_k. Therefore, $S \cap U_k$ is a countable union of slices of U_k.

With that in mind, the path starts from a single slice of U_0, and from there we have countably many options in U_{i_1} and then in U_{i_2} and so on till U_i. In other words, over all possible finite sequences $U_0, U_{i+1}, \ldots, U_{i_r}, U_i$, we can reach only countably many slices of U_i, therefore $K \cap U_i$ is a countable union of slices.

So now K is second countable: there are countably many slices from each U_i and each such slice is second countable and there are countably many U_is and because this particular sequence of operations is nice enough, our infinities remain countable (to put it simply, countably infinity is closed under addition).

Thus, $K \hookrightarrow M$ is a connected integral manifold for $\mathscr{D}$ through p. It is maximal because if $N \xrightarrow{\psi} M$ is another connected integral manifold through p, then by taking the pushforward of piecewise smooth paths in N we can see that $\psi(N) \subseteq K$.

Uniqueness: Let (N, ψ) be another maximal integral submanifold through p, then as observed above, using pushforward of paths, ψ factors through $i\colon K \hookrightarrow M$ via some map $\psi_0\colon N \to K$. From the result above, ψ_0 is smooth. It is injective because ψ is and an immersion because ψ is an immersion. By maximality of N, ψ must be a surjection as well and the dimensions of N, K are equal. As a corollary of the inverse function theorem, ψ_0 is a diffeomorphism. ☐

Remark. Above we have a coherent topology on K generated by the slices. Since each slice is second countable, the moment we know that there are countably many slices we can conclude that K is also second countable by taking the union of countable bases of the individual slices. This of course breaks down when we have an uncountable number of generators, as can be seen with the discrete topology on any uncountable space.

7 A note on complete vector fields

Let M be a smooth manifold and X a complete vector field, i.e., for each $p \in M$, the integral flow through p is defined for all time.

Theorem 22. *We have a smooth map $\Phi\colon M \times \mathbb{R} \to M$ such that $\Phi(p, 0) = p \,\forall\, p \in M, t \mapsto \Phi(p, t)$ is the integral flow through p.*

Proof. The existence of Φ is clear by the existence of maximal integral flows through points. Moreover, by the smooth dependence on initial point, for each $p \in M$ there is a neighbourhood $U \times (-\epsilon, \epsilon)$

of $(p, 0)$ where Φ is smooth. Secondly, from an earlier result, we know that $\Phi_t \colon m \mapsto \Phi(m, t)$ is a diffeomorphism of M. From the same result (and completeness) $\forall\, p, s, t\, \Phi(p, t + s) = \Phi(\Phi(p, t), s)$.

So, given (p, t_0), let $U \times (-\epsilon, \epsilon)$ be a neighbourhood of $(\Phi_t(p), 0)$ where Φ is smooth. Let $V = \Phi_{-t_0}(U)$, which is a neighbourhood of p. Then

$$\begin{aligned} \Phi \colon V \times (t_0 - \epsilon, t_0 + \epsilon) &\to M \\ (q, t_0 + a) &\mapsto \Phi(\Phi_{t_0}(q), a) \end{aligned}$$

factors as $\Phi \circ (\Phi_{t_0} \times \{x \mapsto x - t_0\})$, hence is a smooth map as required. □

Now suppose M, N are smooth manifolds and X is a vector field on $M \times N$ which maps to 0 under the projection $M \times N \to N$, i.e., X is tangent to the fibres. This would mean that the flows corresponding to X remain in the fibres of the projection. Therefore, X is complete iff it's restriction to each fibre is complete on the fibre. In this case, we get a smooth map $\Phi \colon M \times N \times \mathbb{R} \to M \times N \xrightarrow{\pi} M$ such that

1. $\Phi(m, n, 0) = m\, \forall\, m, n$
2. $t \mapsto \Phi(m, n, t)$ is the flow through m in the fibre over n for all n, m.

Now, suppose G is a Lie group. It is a theorem [2] that left invariant vector fields on G are complete (the basic idea is to take translates of curves at identity, so every point has some uniform ϵ duration of translation along the vector field). The Lie algebra $\mathfrak{g}$ is a manifold and on $G \times \mathfrak{g}$ we have the vector field

$$\begin{aligned} F \colon G \times \mathfrak{g} &\to TG \times T\mathfrak{g} \\ (g, X) &\mapsto ((g, (l_g)_{*,e} X), (X, 0)) \end{aligned}$$

where $l_g \colon x \mapsto gx$ is multiplication by $g \in G$ from the left and $(l_g)_{*,e}$ it's derivative at the identity e. This map is a vector field and it is smooth because the coodinates are smooth: by the smoothness of left invariant vector fields and the universality of product. This vector field is tangent to the fibres of $G \times \mathfrak{g} \to \mathfrak{g}$ and each of the restrictions to the fibres are complete, so we have a flow

$$\Phi \colon G \times \mathfrak{g} \times \mathbb{R} \to G$$

which satisfies some additional properties because of the Lie structure on G.

1. By left invariance of the vector fields along fibres, $\Phi(g, X, t) = g\Phi(e, X, t)$.
2. By uniqueness of flows, $\Phi(\Phi(e, X, s), X, t) = \Phi(e, X, s + t)$
3. Combining, $\Phi(e, X, s)\Phi(e, X, t) = \Phi(e, X, s + t)$.
4. From the general theory of flows, $\Phi(e, sX, t) = \Phi(e, X, st)$

Combining all of these tells us that it is enough to know $\Phi(e, X, 1)$ for every X. This gives us the Lie group exponential

$$\begin{aligned} \exp \colon \mathfrak{g} &\to G \\ X &\mapsto \Phi(e, X, 1) \end{aligned}$$

which basically gives us the transport of e along X for one unit of time. Once we have the exponential map for Lie groups we can develop a rich theory of Lie groups and Lie algebras.

7.1 A shorter note on smoothness of left invariant vector fields

This part is purely for me to note something down. Let Γ denote the left invariant vector fields on a Lie group G, then we have the evaluation at $e, \Gamma \to \mathfrak{g}$. This is injective by left invariance. It is easy to obtain a left invariant vector field for a given $X \in \mathfrak{g}$, namely by setting $\tilde{X}(g) = (l_g)_{*,e} X$, what's left is to show that this is a smooth vector field. Let m denote the multiplication on G, then $(l_g)_{*,e}X = m_{*,(g,e)}(0, X)$ and we have

$$\begin{tikzcd}
((g,0),(e,v)) \in TG \times TG \arrow[r, "m_*"] & TG \\
g \in G \arrow[u] \arrow[ur, "\tilde{X}"'] &
\end{tikzcd}$$

which shows that $\tilde{X}$ is a composition of smooth maps, hence smooth.

References

[1] W. Hurewicz, *Lectures on Ordinary Differential Equations*

[2] F. Warner, *Foundations of Differential Manifolds and Lie Groups*

[3] John M. Lee, *Introduction to Smooth Manifolds*

[4] Notes by Marco Gualtieri, Vector fields[link]

[5] Notes by Andrea Rincon, Frobenius Theorem[link]