Disentanglement Math Notes

Berkan Ottlik (bto2106), Jay Ram (jcr2211), and Aaron Liss (ajl2277)

January 6, 2023

1 Review of Group Theory

Definition 1.1 (Fields)**.** A field is a set equipped with two binary operations, one called addition and the other called multiplication, denoted in the usual manner, which are both commutative and associative, both have identity elements (the additive identity denotes 0 and the multiplicative identity denoted 1), addition has inverse elements (the inverse of x is denoted $-x$), multiplication has inverses of nonzero elements (the inverse of x denoted $\frac{1}{x}$ or x^{-1}), multiplication distributes over addition, and $0 \neq 1$.

$\mathbb{Z}$ is not a field, but $\mathbb{R}$ is a field!

Definition 1.2 (Rings)**.** A ring is a set equipped with two binary operations, one called addition, and the other called multiplication, denoted in the usual manner, which are both associative, addition is commutative, both have identity elements (the additive identity denotes 0 and the multiplicative identity denoted 1), addition has inverse elements (the inverse of x is denoted $-x$), and multiplication distributes over addition. If multiplication is also commutative, then the ring is called a commutative ring.

$\mathbb{Z}$ is a ring, but $\mathbb{N}$ is not. Integers modulo n are also rings.

Definition 1.3 (Groups)**.** A group is a set equipped with one binary operation that is associative, has an identity element, and has inverse elements. If, furthermore, multiplication is also commutative, then the group is called a commutative group or an Abelian group. Abelian groups can be denoted either additively or multiplicatively, but non Abelian groups are usually denoted multiplicatively.

The set Z of all objects under the action of all group elements is referred to the orbit of $z \in Z$ under the action of the group G.

2 Review of Differential Geometry

Definition 2.1 (Injective Map (one-to-one))**.** A map $f : X \mapsto Y$ is injective if every element in the co-domain maps to at most one element in the domain. In symbols: $\forall x_1, x_2 \in X : f(x_1) = f(x_2) \implies x_1 = x_2$.

Definition 2.2 (Surjective Map (onto))**.** A map $f : X \mapsto Y$ is surjective if every element in the co-domain maps to at least one element in the domain. In symbols: $\forall y \in Y : \exists x \in X : f(x) = y$.

Definition 2.3 (Isometry)**.** Let (X, d_1) and (Y, d_2) be metric spaces. Then (X, d_1) is said to be isometric to (Y, d_2) if there exists a surjective mapping $f : X \mapsto Y$ such that $\forall x_1, x_2 \in X : d_1(x_1, x_2) = d_2(f(x_1), f(x_2))$. This mapping f is said to be an isometry.

Definition 2.4 (Topology)**.** Let X be a non-empty set. A set $\mathcal{T}$ of of subsets of X is said to be a topology on X if,

(i) $X \in \mathcal{T} \wedge \emptyset \in \mathcal{T}$

(ii) The union of any number of sets (finite or infinite) in $\mathcal{T}$ belongs to $\mathcal{T}$

(iii) The intersection of any two sets in $\mathcal{T}$ belongs to $\mathcal{T}$.

Let's have $X = \{a, b, c, d, e, f\}$, then $\mathcal{T}_1 = \{X, \emptyset, \{a\}, \{c, d\}, \{a, c, d\}, \{b, c, d, e, f\}\}$ is a topology on X, but $\mathcal{T}_1 = \{X, \emptyset, \{a\}, \{c, d\}, \{a, c, d\}, \{b, c, d, e\}\}$ is not a valid topology.

All sets in a topology are open.

Metric spaces are an important class of topological spaces.

The topology on the metric space $M = (A, d)$ induced by (the metric) d is defined as the topology $\mathcal{T}$ generated by the basis consisting of the set of all open ϵ-balls in M.

Definition 2.5 (Neighborhood)**.** Let $(X, \mathcal{T})$ be a topological space, N a subset of X and p a point in N. Then N is said to be a neighborhood of the point p if there exists an open set U such that $p \in U \subseteq N$.

Definition 2.6 (Homeomorphism)**.** Let $(X, \mathcal{T}_1)$ and $(Y, \mathcal{T}_2)$ be topological spaces. Then they are said to be homeomorphic if there exists a function $f : X \mapsto Y$ which has the following properties.

(i) f is bijective

(ii) $\forall U \in \mathcal{T}_2 : f^{-1}(U) \in \mathcal{T}_1$

(iii) $\forall V \in \mathcal{T}_1 : f(V) \in \mathcal{T}_2$

We say that the map f is a homeomorphism between $(X, \mathcal{T}_1)$ and $(Y, \mathcal{T}_2)$.

Definition 2.7 (Hausdorff Space)**.** A topological space $(X, \mathcal{T})$ is said to be a Hausdorff space if for each pair of distinct points $a, b \in X$, there exists open sets U, V such that $a \in U,\ b \in V$ and $U \cap V = \emptyset$.

The topology can differentiate between points and sequences converge to unique points.

For any metric space (X, d) and topology $\mathcal{T}$ induced on X by d, the topological space $(X, \mathcal{T})$ is a Hausdorff Space.

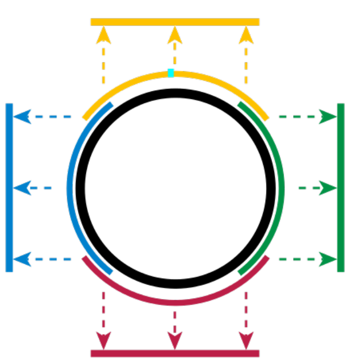
An example of a non-Hausdorff Space is a pseudo metric space, where the distance between two distinct points, A and B, can be zero. Thus any open ϵ-ball containing A also contains B, and any sequence converging to A also converges to B. So this fails the Hausdorff condition.

Definition 2.8 (Locally Euclidean)**.** A topological space $(X, \mathcal{T})$ is said to be locally euclidean if there exists a positive integer n such that each point $x \in X$ has an open neighborhood homeomorphic to an open ball around 0 in $\mathbb{R}^n$.

Definition 2.9 (Topological Manifold)**.** A Hausdorff locally euclidean space is said to be a topological manifold.

Definition 2.10 (Chart)**.** A chart ϕ for a topological space M is a homeomorphism from an open subset W of M to an open subset U of $\mathbb{R}^n$

Definition 2.11 (Atlas)**.** An atlas is a set of charts such that the union of all their domains covers M



The four charts above each map a part W_α of the circle M to an open interval U_α, and the four charts form an atlas of the circle.

Definition 2.12 (Tangent Space)**.** A tangent space to a manifold M of dimension n at point p is the n-dimensional real vector space containing all possible directions in which one can tangentially pass through point p. It is denoted $T_p(M)$. In physics, the tangent space to a manifold at a point is equivalent to the space of possible velocities for a particle moving on the manifold. It is made up of all velocity vectors $\gamma(t)$ where $\gamma : \mathbb{R} \mapsto \ M$ is a path such that $\gamma(t) = p$.

For example, circle M is a 1-dimensional manifold in $\mathbb{R}^2$, and the tangent space at any point p is the set of all vectors on the tangent line at point p.

Definition 2.13 (Differentiable Manifold)**.** A differentiable manifold of dimension n is a set M together with a family of injective maps $\mathbf{x}_\alpha : U_\alpha \mapsto M$ of open sets $U_\alpha \subseteq \mathbb{R}^n$ into M such that,

(i) $\bigcup_\alpha \mathbf{x}_\alpha(U_\alpha) = M$

(ii) $\forall \alpha, \beta$ where $\mathbf{x}_\alpha(U_\alpha) \cap \mathbf{x}_\beta(U_\beta) = W \neq \emptyset$, we have that $\mathbf{x}_\alpha^{-1}(W),\ \mathbf{x}_\beta^{-1}(W)$ are open sets in $\mathbb{R}^n$, and $\mathbf{x}_\alpha^{-1} \circ \mathbf{x}_\beta,\ \mathbf{x}_\beta^{-1} \circ \mathbf{x}_\alpha$ are continuous and differentiable everywhere. Notice that $\mathbf{x}_\alpha^{-1} \circ \mathbf{x}_\beta$ and $\mathbf{x}_\beta^{-1} \circ \mathbf{x}_\alpha$ are $U_\beta \mapsto U_\alpha$ and $U_\alpha \mapsto U_\beta$ respectively.

(iii) The family $\{U_\alpha, \mathbf{x}_\alpha\}$ is maximal relative to conditions 1 and 2. In other words, $\{U_\alpha, \mathbf{x}_\alpha\}$ is not properly contained in any other family satisfying (i) and (ii).

Definition 2.14 (Differentiable Map)**.** Let M and N be differentiable manifolds. A differentiable map is any $f : M \mapsto N$ that is continuous and differentiable everywhere.

From above, $\mathbf{x}_\alpha^{-1} \circ \mathbf{x}_\beta$ and $\mathbf{x}_\beta^{-1} \circ \mathbf{x}_\alpha$ are differentiable maps $x_\beta^{-1}(W) \mapsto x_\alpha^{-1}(W)$ and $x_\alpha^{-1}(W) \mapsto x_\beta^{-1}(W)$ respectively.

Definition 2.15 (Diffeomorphism)**.** Let M and N be differentiable manifolds. A differentiable map $f : M \mapsto N$ is a diffeomorphism if it is a bijection and its inverse is also a differential map.

Whereas a homeomorphism is a bijection that is continuous with a continuous inverse, a diffeomorphism is additionally differentiable and has a differentiable inverse.

From above, $\mathbf{x}_\alpha^{-1} \circ \mathbf{x}_\beta$ and $\mathbf{x}_\beta^{-1} \circ \mathbf{x}_\alpha$ are diffeomorphisms .

Definition 2.16 (Cotangent Space)**.** The cotangent space to a differentiable manifold M of dimension n at point p is the n-dimensional real vector space containing all gradients of differentiable functions at x, and for finite dimensional manifolds is the dual space to the tangent space.

Definition 2.17 (Riemannian Manifold)**.** A Riemannian manifold is an n-dimensional differentiable manifold M together with a choice, for each $p \in M$, of an inner product $\langle\ ,\ \rangle$ in $T_p(M)$ that varies differentiably with p in the following sense. For some (hence, all) parameterization $\mathbf{x}_\alpha : U_\alpha \mapsto M$ with $p \in \mathbf{x}_\alpha(U_\alpha)$, the functions

$$g_{ij}(u_1, \dots, u_n) = \langle \frac{\partial}{\partial u_i}, \frac{\partial}{\partial u_j} \rangle,\ i, j = 1, \dots, n$$

are differentiable at $x_\alpha^{-1}(p)$; here $(u_1, \dots, u_n)$ are the coordinates of $U_\alpha \subseteq \mathbb{R}^n$.

Therefore we define distances locally in our manifold. If we choose a basis for the tangent space $T_p(M)$, then in that basis we can represent our metric as a positive definite matrix $G_p \in \mathbb{S}_+^n$ and $\langle \mathbf{v}, \mathbf{w} \rangle_p = \mathbf{v}^\top G_p \mathbf{w}$. The l_2 metric, for example, is the special case $G_p = \mathbb{I}$. For cotangent vectors, the metric is $\langle \mathbf{v}, \mathbf{w} \rangle_p^* = \mathbf{v}^\top G_p^{-1} \mathbf{w}$. This metric can change across the manifold though! The geodesic distance between any two points is the minimum length of any path between them. The geodesic is the locally shortest path that is parameterized by arc length (this isn't necessarily the shortest path). Geodesic distance is global distance, and the metric is local distance. You need to have a Riemannian Manifold to define distances because you need the metric.

Quantities on the manifold must be defined in a way that they transform consistently between different embeddings, since there are many different ways that a manifold can be embedded in a vector space (for instance, the manifold of natural images can be embedded in the vector space of pixel representations of an image). Let $p \in \mathbb{R}^n$ be an embedding of the point p. Under a differentiable change in embedding $\bar{p} = f(p)$, tangent vector components v transform as $\bar{v} = \mathbb{J}_f v$ where $\mathbb{J}_f$ is the Jacobian of f at p. Cotangent vector components w transform as $\bar{w} = \mathbb{J}_f^{-1} w$. A linear transform of vectors in $T_p(M)$ represented by the matrix A transforms as $\bar{A} = \mathbb{J}_f A \mathbb{J}_f^{-1}$. The metric matrix $G_p \in \mathbb{S}_+^n$ transforms as $G_{\bar{p}} = \mathbb{J}_f^{-T} G_p \mathbb{J}_f^{-1}$.

Definition 2.18 (Laplacian Operator)**.** The Laplacian $\triangle f$ of a function f measures the divergence of a the function's gradient: $\triangle f = \nabla^2 f$, where $\nabla = (\frac{\partial}{\partial x_1}, ..., \frac{\partial}{\partial x_n})$. It can be expressed explicitly as the sum of the pure second partial derivatives with respect to each vector of an orthonormal basis for $\mathbb{R}^n$: $\triangle f = \sum_i \frac{\partial^2 f}{\partial x_i^2}$

Definition 2.19 (Laplace-Beltrami Operator)**.** The Laplace–Beltrami operator is a generalization of the Laplace operator to functions defined on Riemannian manifolds. Once the metric G_p is known in a given coordinate system, the Laplace-Beltrami Operator can be constructed in terms of the coordinates: $\triangle[f](x) = \frac{1}{\sqrt{det(G_p)}} \sum\limits_j \frac{\partial}{\partial x_i} (\sqrt{det(G_p)} \sum\limits_i g_{ij}^{-1} \frac{\partial f}{\partial x_i})$. In flat Euclidean space, $G_p = I$, which reduces this to the familiar $\sum_i \frac{\partial^2 f}{\partial x_i^2}$.

Definition 2.20 (Tangent Bundle)**.** Let S be some abstract surface, then we can define $T(S) = \{(p, w) | p \in S,\ w \in T_p(S)\}$ to be the tangent bundle of S.

Definition 2.21 (Parametrized Curve)**.** A parametrized curve $\alpha : [0, l] \mapsto S$ is the restriction to $[0, l]$ of a differentiable mapping of $(0 - \epsilon, l + \epsilon)$, $\alpha > 0$, into S. If $\alpha(0) = p$ and $\alpha(l) = q$, we say that α joins p to q. α is regular if $\alpha'(t) \neq 0$ for $t \in [0, l]$.

$I = [0, l]$ whenever specification of the endpoint l is unnecessary.

Definition 2.22 (Geodesic)**.** A nonconstant, parametric curve $\gamma : I \mapsto S$ is geodesic at $t \in I$ if the field of its tangent vectors $\gamma'(t)$ is parallel along γ at t. γ is a parametrized geodesic (also just called a geodesic) if it is geodesic for all $t \in I$.

The geodesic between x and y is the shortest path parameterized by arc length. It isn't necessarily a minimum path from start to end. For example, the great circle from the north pole to itself is a geodesic, even though the distance of the shortest path is zero.

Definition 2.23 (Geodesic Distance)**.** For any two points on a manifold, $x, y \in M$, the geodesic distance between them is defined as the minimum length of any path between them:

$$\mathcal{D}(x, y) = \min_{\substack{\gamma \\ \gamma(0)=x \\ \gamma(1)=y}} \int_0^1 dt \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}}$$

Whereas the metric is a local notion of distance, defined only in the tangent space, the geodesic distance is a global distance between two points on a manifold.

In machine learning, "metric learning", despite the name, typically refers to learning a single, global notion of distance (the geodesic), or to learning a mapping that preserves distances, under the assumption that the correct local distance (the metric) is already known.

Definition 2.24 (Geodesically Complete)**.** Also called a complete manifold, a Riemannian manifold M is geodesically complete if every geodesic $\gamma : I \mapsto M$ is maximal, such that $I = (-\infty, \infty)$.

Informally, a Riemannian manifold is geodesically complete if at any point p you can follow a "straight" line indefinately along any direction.

Definition 2.25 (Simply Connected)**.** A manifold is simply connected if any closed loop can be continuously deformed into a single point.

Definition 2.26 (Vector Field)**.** A vector field w along a paremeterized curve $\alpha : I \mapsto S$ is a correspondence that assigns each $t \in I$ a vector $w(t) \in T_{\alpha(t)}(S)$

Definition 2.27 (Parallel Vector Field)**.** A vector field w along a parameterized curve $\alpha : I \mapsto S$ is said to be parallel if $\frac{Dw}{dt} = 0$ for every $t \in I$.

Proposition 2.1 (Constants among parallel vector fields). If w and v are parallel vector fields along $\alpha : I \mapsto S$, then the following are all constant: $\langle w(t), v(t) \rangle$, $|w(t)|$, $|v(t)|$, and the angle between $w(t)$ and $v(t)$.

Proposition 2.2 (Uniqueness of parallel vector fields). Let $\alpha : I \mapsto S$ be a parameterized curve in S and let $w_0 \in T_{\alpha(t_0)}(S),\ t_0 \in I$. Then there exists a unique parallel vector field $w(t)$ along $\alpha(t)$ with $w(t_0) = w_0$.

How do we relate a vector in one tangent space to a vector in another tangent space? In general, there is no unique mapping from vectors in one tangent space to another. A vector in a tangent space $T_x(M)$ can be identified with a vector in $T_y(M)$ in a path-dependent manner using parallel transport! In parallel transport, a vector is moved infinitesimally along a path such that it is always locally parallel with itself as it moves.

Definition 2.28 (Parallel Transport)**.** Let $\alpha : I \mapsto S$ be a parameterized curve and $w_0 \in T_{\alpha(t_0)}(S), t_0 \in I$. Let w be the parallel vector field along α, with $w(t_0) = w_0$. The vector $w(t_1), t_1 \in I$ is called the parallel transport of w_0 along α at the point t_1.

Definition 2.29 (Affine Connection)**.** The affine connection at p is a map $\Gamma_p : T_p(M) \times T_p(M) \mapsto T_p(M)$. For vectors $\mathbf{v}, \mathbf{w} \in T_p(M), \Gamma_p(\mathbf{v}, \mathbf{w})$ can be thought of as the amount that $\mathbf{v}$ changes when moving to a nearby tangent space in the direction $\mathbf{w}$.

For a Riemannian manifold, the affine connection should preserve the metric (meaning that the inner product between vectors does not change as they are parallel transported), and it should be torsion free (meaning the vectors should not "twist" as they are parallel transported. The Levi-Civita connection is the unique affine connection that preserves these properties for a Riemannian manifold.

Definition 2.30 (Levi-Civita Connection)**.** This is an affine connection that preserves the metric (meaning inner products don't change) and is torsion-free (meaning vectors don't twist during transport). For a given choice of coordinates such that the metric can be represented by G_p at p, and letting the ijth element of G_x be denoted g_{ij}, and the ith element of $\Gamma_p(\mathbf{v}, \mathbf{w})$ be denoted as $\Gamma_p(\mathbf{v}, \mathbf{w})_i$, the Levi-Civita connection at p can be written as.

$$\Gamma_p(\mathbf{v}, \mathbf{w})_i = \sum_{jk} \Gamma^i_{jk} v_j w_k$$

$$\Gamma^i_{jk} = \frac{1}{2} \sum_l g^{-1}_{il} \left(\frac{\partial g_{lk}}{\partial p_j} + \frac{\partial g_{lj}}{\partial p_k} - \frac{\partial g_{jk}}{\partial p_l} \right)$$

This defines the covariant derivative, which is essentially the projection of the ordinary derivative onto the manifold for a manifold embedded in $\mathbb{R}^n$, but in general is thought of a correction to force the covariant derivative to transform correctly as a rank-(0,1) tensor.

Definition 2.31 (Holonomy Group)**.** Let S be a regular surface and $p \in S$. For each piecewise regular parameterized curve $\alpha : [0, l] \mapsto S$ with $\alpha(0) = \alpha(l) = p$, let $P_\alpha : T_p(S) \mapsto T_p(S)$ be the map which assigns to each $v \in T_p(S)$ its parallel transport along α back to p. By proposition 2.1, P_α is a linear isometry of $T_p(S)$. If $\beta : [l, \bar{l}]$ is another piecewise regular parameterized curve with $\beta(l) = \beta(\bar{l}) = p$, define the curve $\beta \circ \alpha : [0, \bar{l}] \to S$ by running successively first α and then β; that is $\beta \circ \alpha(s) = \alpha(s)$ if $s \in [0, l]$ and $\beta \circ \alpha(s) = \beta(s)$ if $s \in [l, \bar{l}]$.

In the paper, they describe a homology more simply as a linear transformation H_γ of the loop $\gamma : [0, 1] \mapsto \mathbb{M}$ where $\gamma(0) = \gamma(1) = p$. The columns of this matrix are specified by an orthonormal basis $\{\mathbf{e}_i\} \subseteq T_p(M)$ of the tangent space at x. H_γ characterizes how any vector transforms when parallel transported around a loop. The space of all holonomies for all loops that start and end at point p for a given Riemannian manifold is the holonomy group, denoted as $\text{Hol}_p(M)$.

Proposition 2.3 (Invariant Subspaces Under $\text{Hol}_p(M)$ Group Action). If manifold M is a product of submanifolds $M_1 \times M_2 \times \ldots \times M_n$, with the corresponding product metric as its metric, then the tangent space $T_p(M)$ can be decomposed into orthogonal subspaces $T_p^{(1)}(M), \ldots, T_p^{(n)}(M)$ such that the action of $\text{Hol}_p(M)$ leaves each subspace invariant. That is, if $v \in T_p^{(i)}(M)$, then $H_\gamma v \in T_p^{(i)}(M)$ for all γ. The converse also holds locally, and it holds globally too if M is simply connected and geodesically complete.

The $T_p^{(i)}(M)$ subspaces are each tangent to the respective submanifolds that make up M.

Definition 2.32 (Lie Group)**.** Groups that are also smooth manifolds, such as rotations.

More formally, a real Lie group is a group that is also a finite-dimensional real smooth manifold, in which the group operations of multiplication and inversion are smooth maps. Smoothness of the group multiplication means that the following map μ is a smooth mapping of the product manifold into the original manifold.

$$\mu : M \times M \mapsto M, \ \mu(x, y) = xy$$

Definition 2.33 (de Rham Decomposition Theorem)**.** Given M, a simply connected and complete Riemannian manifold, there exists a unique decomposition up to isometry and permutation of factors

$$M = \Pi_{i=1}^n M_i$$

Where M_i are complete, simply connected Riemannian irreducible manifolds. Moreover the holonomy representation of M over $T_x(M)$ is the product of holonomy representations of M_i over $T_{x_i}(M_i)$ where $x = (x_1, \ldots, x_n)$.