A Background

A.1 Riemannian geometry

Riemannian geometry is a part of differential geometry that studies Riemannian manifolds, smooth manifolds equipped with a Riemannian metric. A manifold $\mathcal{M}$ is a generalization and abstraction of the notion of a curved surface. It is a topological set that is modeled closely on Euclidean space locally but may vary widely in global properties. It means for each $p \in \mathcal{M}$, one can associate a tangent space $T_p(\mathcal{M}) \subseteq \mathbb{R}^d$, corresponding to the union of all tangent vectors of differentiable curves passing through p. Consequently, at any point $p \in \mathcal{M}$, the tangent space $T_p(\mathcal{M})$ is spanned by a basis $\{\partial_{x_i,p}\}_{i=1}^N$:

$$T_p(\mathcal{M}) = \{\dot{\gamma}(0) | \gamma : \mathbb{R} \to \mathcal{M} \in \mathcal{C}^1 \text{and } \gamma(0) = p\} = \text{span}\{\partial_{x_i,p}\}_{i=1}^N. \tag{5}$$

A Riemannian manifold is a differentiable manifold $\mathcal{M}$ with a Riemannian metric $\mathcal{R}$, a 2-tensor field, such that at each $p \in \mathcal{M}$, we have a function $\mathcal{R}|_p : T_p(\mathcal{M}) \times T_p(\mathcal{M}) \to \mathbb{R}$ which is symmetric and positive definite. At each $g \in G$, the Riemannian metric can be expressed by its local representation, a symmetric positive definite matrix $\boldsymbol{R}_p$ whose components are given by:

$$r_{ij}(p) = \mathcal{R}|_p(\partial_{x_i,p}, \partial_{x_j,p}). \tag{6}$$

In the following we consider Riemannian metric with diagonal local representation at the origin e, that is:

$$r_{ij}(e) = \begin{cases} r_i & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}. \tag{7}$$

The Riemannian metric induces an inner product, such that for any $u = u^i \partial_{x_i,e}$ and $v = v^i \partial_{x_i,e}$, we have $\langle u, v \rangle_{\mathcal{R}(e)} = u^i r_i v^i$ and $||u||^2_{\mathcal{R}(e)} = u^i r_i u^i$. Between each pair of points p and q of a Riemannian manifold, we define their Riemannian distance as the length of the shortest curve connecting the two points (a.k.a. geodesic). More formally, we have:

$$d(p, q) = \inf_{\gamma \in \mathcal{C}^\infty_{p,q}} \int_0^1 ||\dot{\gamma}(\tau)||_{\mathcal{R}_{\gamma(\tau)}} d\tau, \tag{8}$$

where $\mathcal{C}^\infty_{p,q} = \{\gamma : [0, 1] \to \mathcal{M} | \gamma \in \mathcal{C}^\infty, \gamma(0) = p, \gamma(1) = q\}$. An more convenient way to define the Riemannian between two points is via the exponential and logarithmic Riemannian maps (Figure 5). The Riemannian exponential map $\exp_p T_p(\mathcal{M}) \to \mathcal{M}$ is defined as:

$$\exp_p(v) = \gamma_{p,v}(1), \qquad (\text{and by extension } \exp_p(tv) = \gamma_{p,v}(t)), \tag{9}$$

where $\gamma_{p,v}(1)$ is the unique Riemannian geodesic starting at $p \in \mathcal{M}$ with initial velocity $v \in T_p(\mathcal{M})$. We can interpret the exponential map as follows. Let's choose any direction in our tangent space and follow it with a step forward. We make sure to take the shortest path and end up at a new point. This process is what we called the exponential map.[6] The inverse mapping of the exponential map, the logarithmic map $\log_p : \mathcal{M} \to T_p(\mathcal{M})$ will map an element $p \in \mathcal{M}$ to the smallest vector $v = \log_p q \in T_p(\mathcal{M})$ as measured by the Riemannian metric such that $q = \exp_p v \in \mathcal{M}$.

A.2 Group theory

A group $(G, \cdot)$ is a set G equipped with a binary operation $\cdot : G \to G$ called group product, satisfying the four group axioms (closure, associativity, identity element and inverse elements). To map the structure of the group to some mathematical object, one requires a representation. We define H as the vector space to which our mathematical object belongs and $\mathcal{B}(H)$ the space of bounded linear invertible operators $H \to H$. A representation $\mathcal{V} : G \to \mathcal{B}(H)$ maps a group element to an operator such that the identity element, the group product and the group inverse are preserved. We define the left-regular representation $\mathcal{L}_g$ on the (infinite-dimensional) vector space of functions $G \to \mathbb{R}^d$ via:

$$(\mathcal{L}_g \circ f)(h) = f(g^{-1}h), \tag{10}$$

with $f : G \to \mathbb{R}^d$ a function on the group G and g, h elements of the group G. Using this group representation, we can formally define the left-invariance (a.k.a. equivariance).

[6] It comes from the fact that all these tiny steps magically resemble the series expansion of the exponential function.

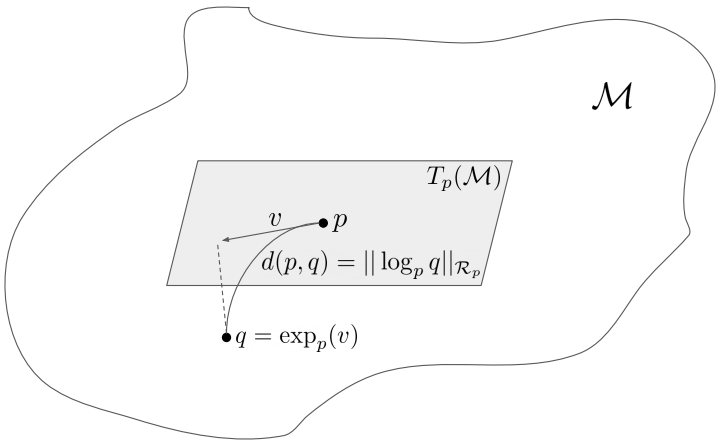


Figure 5: Riemannian exponential map $\exp_p : T_p(\mathcal{M}) \to \mathcal{M}$ and Riemannian logarithmic map $\log_p : \mathcal{M} \to T_p(\mathcal{M})$.

Definition A.1 (Equivariance) *An operator $\Phi : \mathcal{X} \to \mathcal{Y}$ from one vector space to the other is equivariant (or left-invariant under group transformation) if it satisfies the following property:*

$$\mathcal{L}'_g \circ \Phi = \Phi \circ \mathcal{L}_g, \qquad \forall g \in G. \tag{11}$$

Equivariance can be realized in many ways, and in particular, the group representations $\mathcal{L}_g$ and $\mathcal{L}'_g$ need not be the same, as they act on different spaces $\mathcal{X}$ and $\mathcal{Y}$. Note that the familiar concept of invariance is a special kind of equivariance where $\mathcal{L}'_g$ is the identity transformation for all element g of the group G.

A Lie group is a continuous group whose group elements are parameterized by a finite-dimensional differentiable manifold. In essence, this means that a Lie group is a group to which we can apply differential geometry. From now on, we also assume the group manifold is equipped with a Riemannian metric, that is it is Riemannian manifold. To each element g of G, we can associate a tangent space $T_g(G)$, which is spanned by a basis of left-invariant vectors denoted by $\{\mathcal{A}_i|_g\}_{i=1}^n$. We write $T_g(G) = \text{span}\{\mathcal{A}_1|_g, \ldots, \mathcal{A}_n|_g\}$ and we can express tangent vectors $\dot{\gamma}(t)$ of $T_g(G)$ in this basis via $\dot{\gamma}(t) = c^i(t)\mathcal{A}_i|_{\gamma(t)}$.

Moreover, because this frame of basis vector is left-invariant, the coefficients $c_i(t)$ remain unchanged if the curve is moved by applying left group product. The tangent space at the origin $T_e(G)$ is spanned by a basis $\{A_i\}_{i=1}^n$ where $A_i = \mathcal{A}_i|_e$. We make a subtle difference in notation between $\mathcal{A}$ and A, where $\mathcal{A}$ represents a whole vector field and $\mathcal{A}|_g$ represents the vector at location g. The straight A is used to indicate a vector in the Lie algebra, the tangent space at the origin. Via the push-forward $(L_g)_*$, we can generate a whole vector space by picking a vector in the Lie algebra and transporting tangent vectors from $\dot{\gamma}(0) \in T_e(G)$ to $g \cdot \dot{\gamma}(0) \in T_g(G)$ using:

$$\mathcal{A}_i|_g = (L_g)_* A_i, \qquad \forall g \in G. \tag{12}$$

The left-invariant frame of basis vectors could also be interpreted as the directional derivative of functions defined on the group G. At any element g, the value of the directional derivative of a function f defined on G can be computed using the push-forward operation:

$$\mathcal{A}_i|_g f = (L_g)_* \mathcal{A}_i|_e f := \mathcal{A}_i|_e (f \circ L_{g^{-1}}), \tag{13}$$

so $\mathcal{A}_i|gf$ represents the directional derivative along the vector field at location g, which can be defined by translating the function back to the origin via $L_{g^{-1}}$ and compute the derivative at the origin along the direction specified by $\mathcal{A}_i|_e$. In the above $L_g h := gh$.

The Laplace-Beltrami operator, the generalization to Riemannian manifold of the Laplace operator[7] is defined below.

Definition A.2 (Laplace-Beltrami operator) *Let g be an element of the Lie group $(\mathcal{M}, \cdot)$. The Laplace-Beltrami operator on the Riemannian manifold $\mathcal{M}$ is defined as*

$$\Delta_{\mathcal{R}(g)} = \operatorname{div}(\nabla_{\mathcal{R}(g)}) \tag{14}$$

Using the left-invariant frame of basis vectors as directional differential operators, we can define the gradient, expressed as a vector relative to the basis $\mathcal{A}_i|_g$, of a function $f : G \to G$:

$$\nabla_{\mathcal{R}(g)} f = \mathbf{R}_g^{-1}(\mathcal{A}_1|_g f, \ldots, \mathcal{A}_n|_g f)^\top, \tag{15}$$

where $\mathbf{R}_g$ is the Riemannian metric tensor defined relative to the basis $\mathcal{A}_i$, that corrects for local scaling and shrinking of the manifold as measured by the metric tensor. The divergence of a vector field $F : G \to \mathbb{R}^n$ is the operator:

$$\operatorname{div}(F) = \sum_{i=1}^{n} \mathcal{A}_i|_g F_i. \tag{16}$$

The Laplace-Beltrami operator depends on a Riemannian metric tensor, which describes how lengths of vectors should be measured in different directions, and in the Laplace-Beltrami operator, it rescales the derivates accordingly. While the usual Laplacian is isotropic (derivatives are threated the same in each direction), the Laplace-Beltrami operator can be anisotropic due to the Riemannian metric that is used.

Theorem A.1 (Left-invariance of the Laplace-Beltrami operator) *The Laplace-Beltrami operator $\Delta_{\mathcal{R}(g)}$ is left-invariant and satisfies:*

$$\Delta_{\mathcal{R}(g)} = (L_g)_* \Delta_{\mathcal{R}(e)}. \tag{17}$$

A Lie algebra $\mathfrak{g}$ is a vector space (here the tangent space at the identity element $T_e(G)$) that is endowed with a binary operator called the Lie bracket or commutator $[\cdot, \cdot] : T_e(G) \times T_e(G) \to T_e(G)$ that is bilinear, alternative and satisfies the Jacobi identity. Conceptually, the Lie bracket $[A, B]$ of two vector fields A, B is the derivative of B along the flow generated by A.[8] In the following, we take $\{A_i\}_{i=1}^n$ with $A_i = \mathcal{A}_i|_e$ as Lie Algebra and get left-invariant vector fields via the push-forward operation.

The Lie group exponential and logarithmic maps define the mappings between the group and the tangent space. The exponential map on a Lie group can be thought of as picking a vector A in the Lie algebra, construct a left-invariant vector field $\mathcal{A}$ via the push-forward operator, and follow this vector field by taking infinitesimal steps along the direction indicated by the vector field. Integrating along the vector field defined by A for unit time brings to some point $g \in G$. So $g = \exp A$, where $\exp : \mathfrak{g} \to G$.[9]

[7] Another equivalent definition in local coordinates exists too. It explicitly takes into account the Riemannian metric tensor field $\mathcal{R}$ previously defined. Here we use a more compact definition but using the subscript $\mathcal{R}(g)$ to indicate that the Laplace-Beltrami operator depends on the location on the Riemannian manifold and the Riemannian metric tensor field.

[8] Two vector fields are commutative (zero Lie bracket) if and only if its flows are too, in the sense that there is no difference starting at one point p, traveling a time t_a over the flow of A and then a time t_b over the flow of B, or, instead, traveling first t_b over the flow of B and then t_a over the flow of A.

[9] Note that Riemannian and Lie group exponential maps are different. For the general Riemannian exponential map, the vector field along which we compute the path integral is defined by the Riemannian metric. In the Lie group exponential map, the vector field is defined by the push forward of left-multiplication. The curves defined by the path integrals in the Riemannian exponential are known to be geodesics. The exponential curves in the Lie group case are "straight curves" with respect to a moving frame of reference but are not necessarily geodesics. The Riemannian and Lie group exponential maps only coincide when the Riemannian metric is both left and right invariant (see Section 4.5 in Bekkers [2017]).

A.3 Graph theory

Graphs are generic data representation forms that are useful for describing the geometric structure of data domains [West et al., 1996]. More formally, we define a graph $\mathcal{G}$ as a structure modeling a finite set of interactions called edges $\mathcal{E}$ between a finite set of objects called vertices $\mathcal{V}$. In the following, we denote by $|\mathcal{V}|$ the number of vertices and by $|\mathcal{E}|$ the number of edges. We denote v_i the i-th vertex and $e(v_i, v_j)$ the potential edge from v_i to v_j. We call the neighborhood of the vertex v_i the set of vertices connected to v_i by an edge and denote it by $\mathcal{N}(v_i)$. More generally, we write $\mathcal{N}^k(v_i)$ for the k-hops neighborhood of vertex v_i, that is, the set of vertices connected to v_i with a path of at most k edges. In some cases, it can be useful to add weights on graph edges. In general, the weights can take any value. Nevertheless, in this thesis, we assume weights in the range $[0, 1)$ and measuring the similarity rather than the distance. When the edges' weights are not naturally defined by an application, a common way to define them is to apply a kernel $\mathfrak{K} : \mathbb{R}_+ \to [0, 1)$ on the distance between connected vertices. For theoretical convergence results, we use a Gaussian weighting scheme:

$$w(v_i, v_j) = \begin{cases} \exp\left(-\frac{d^2(v_i,v_j)}{4t}\right) & \text{if } e(v_i, v_j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}, \tag{18}$$

where $d(v_i, v_j)$ denotes the (Riemannian) distance between vertices v_i and v_j and t is a positive real number called bandwidth of the Gaussian kernel. From now on, we furthermore assume that the graphs are undirected and without self-loop.

The field of signal processing on graphs merges algebraic and spectral graph theoretic concepts with computational harmonic analysis to process such signals on graphs [Shuman et al., 2013]. A signal on a graph is a function $f : \mathcal{V} \to \mathbb{R}^d$, mapping each vertex of the graph to a d-dimensional real valued vector. In matrix form, this signal is a $|\mathcal{V}| \times d$ real valued matrix $\boldsymbol{f}$ whose rows are given by $\boldsymbol{f}_i = f(v_i) \in \mathbb{R}^d$.

An essential object in graph signal processing is the Laplacian operator. Under some specific conditions that we will state later, it can be interpreted as a discrete version of the Laplace-Beltrami operator.

Definition A.3 (Symmetric normalized Laplacian) *Let $\mathcal{G}$ be an undirected weighted graph without self-loops. The symmetric normalized Laplacian $\boldsymbol{\Delta}$ is the $|\mathcal{V}| \times |\mathcal{V}|$ real valued matrix whose components are given by:*

$$\Delta_{i,j} = \begin{cases} 1 & \textit{if } i = j \textit{ and } \deg(v_i) > 0 \\ -\frac{w(v_i,v_j)}{\sqrt{\deg(v_i)\deg(v_j)}} & \textit{if } i \neq j \textit{ and } \deg(v_i) > 0 \\ 0 & \textit{otherwise} \end{cases}. \tag{19}$$

Assuming a function f defined on the graph vertices $\mathcal{V}$, by inspection on the components of $\boldsymbol{\Delta}^k \boldsymbol{f}$, we remark that the Laplacian acts as a k-op neighborhood operator. We can prove that $\boldsymbol{\Delta}$ is a symmetric positive definite matrix. Hence it admits a unique eigendecomposition of the form $\boldsymbol{\Delta} = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^\top$ where the j-th column of Φ correponds to the eigenvector ϕ_j associated with real positive eigenvalue λ_j. In these settings, the eigenvalues are in the range $[0, 2]$ [Chung and Graham, 1997]. By analogy with the Euclidean case where the Laplacian's eigenfunctions correspond to the Fourier basis, we can construct a graph Fourier basis from the eigendecomposition of the graph Laplacian, and define the graph Fourier transform and its inverse.[10]

Definition A.4 (Graph Fourier transform) *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph with Laplacian $\boldsymbol{\Delta}$ and let $f : \mathcal{V} \to \mathbb{R}$ be a signal defined on the graph's vertices. The graph Fourier transform $\hat{f}$ of f is given by:*

$$\hat{f}(\lambda_j) = \hat{f}_j = \{\boldsymbol{\Phi}^\top \boldsymbol{f}\}_j = \sum_i \phi_{ij} f_i = \sum_{v_i} \phi(v_i, \lambda_j) f(v_i), \tag{20}$$

and its inverse transform by:

$$f(v_i) = f_i = \{\boldsymbol{\Phi}\hat{\boldsymbol{f}}\}_i = \sum_j \phi_{ij} \hat{f}_j = \sum_{\lambda_j} \phi(v_i, \lambda_j) \hat{f}(\lambda_j). \tag{21}$$

[10] The existence of an inverse transform is a direct consequence of the orthonormality of the eigenvectors.

B Lie groups

At the moment, all the group's elements we are interested in are in 1-1 correspondence with elements of the 3-dimensional general linear group $GL(3)$, the group of 3×3 real matrices. We will use this representation since working with matrices is more convenient. Indeed, in this case the group product, group inverse, group exponential and group logarithm respectively coincide with the matrix product, matrix inverse, matrix exponential and matrix logarithm.

For the roto-translation group $SE(2)$, the spatial part corresponds to the planar translations and the orientation part to the rotation angles. Group elements $g \in SE(2)$ are given in matrix formulation by:

$$g = (x, y, \theta) \leftrightarrow \boldsymbol{G_g} = \begin{pmatrix} \cos\theta & -\sin\theta & x \\ \sin\theta & \cos\theta & y \\ 0 & 0 & 1 \end{pmatrix}, \tag{22}$$

hence the group is 3 dimensional with two parameters x, y coming from the spatial space $\mathbb{R}^2$ and one orientation parameter coming from $[-\pi, \pi)$.

The Lie algebra $\mathfrak{se}(2)$ is the set of 3×3 matrices:

$$\boldsymbol{A_1} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{A_2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \text{ and } \quad \boldsymbol{A_3} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{23}$$

Naturally, it is impossible to uniformly sample elements on this group because the $\mathbb{R}^2$ space is infinite. Adding boundaries to the euclidean space, we get the $[0, 1)^2$ space, which is not homogeneous anymore, but on which it is possible to uniformly sample $|\mathcal{V}_s|$ elements. Because we are considering anisotropic Laplace-Beltrami operators which are symmetric under reflections, it is sufficient to sample the rotation angles θ in the range $[-\pi/2, \pi/2)$.

As defined by its Lie algebra, the matrix logarithm related to an element of the $SE(2)$ group has the form:

$$\log_e \boldsymbol{G}_g = \begin{pmatrix} 0 & -c_3 & c_1 \\ c_3 & 0 & c_2 \\ 0 & 0 & 0 \end{pmatrix} \tag{24}$$

It results that the logarithmic map $\log_e : SE(2) \to \mathfrak{se}(2)$ admits a closed form expression $\log_e g = (c_1, c_2, c_3)^\top$ with[11]:

$$c_1 = \frac{\theta}{2}\left(y + x\cot\frac{\theta}{2}\right), \quad c_2 = \frac{\theta}{2}\left(-x + y\cot\frac{\theta}{2}\right), \text{ and } \quad c_3 = \theta. \tag{25}$$

The group $SO(3)$ of all rotations about the origin of 3-dimensional Euclidean space can be split into a "spatial" part which is the sphere, and a rotation part, which is a rotation around a particular reference axis. Using ZYZ representation, group elements $g \in SO(3)$ are given in matrix formulation by:

$$g = (\alpha, \beta, \gamma) \leftrightarrow \boldsymbol{G_g} = \boldsymbol{R}_{\gamma,z}\boldsymbol{R}_{\beta,y}\boldsymbol{R}_{\alpha,z}, \tag{26}$$

where $\alpha \in [-\pi, \pi]$, $\beta \in [-\pi/2, \pi/2]$ and $\gamma \in [-\pi, \pi]$. We view the sphere S^2 as the spatial part, just like we view our Earth as locally flat. S^2 is parametrized with Euler angles β and γ, which are independent of α. The rotation part is then parametrized by α.

The Lie algebra $\mathfrak{so}(3)$ is the set of antisymmetric 3×3 matrices:

$$\boldsymbol{A_1} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{A_2} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \text{ and } \quad \boldsymbol{A_3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}. \tag{27}$$

To uniformly sample on the group $SO(3)$ is quite challenging because a perfectly uniform sampling on the sphere does not exist. Fortunately, this task has been largely studied, and many algorithms

[11] In the isotropic case, corresponding to the 2-d Euclidean space, one can check that using the exact same expression of the logarithmic map with $\theta = 0$ gives the Euclidean distance.

have been proposed: equiangular[12] [Driscoll and Healy, 1994], HEALPix [Gorski et al., 2005], and icosahedral [Baumgardner and Frederickson, 1985] samplings. As before, due to the same symmetry argument, it is sufficient to restrict the orientation range to $[-\pi/2, \pi/2)$ and to sample $|\mathcal{V}_o|$ elements on this restricted range.

The matrix logarithm related to an element of the $SO(3)$ group has the form:

$$\log_e \boldsymbol{G}_g = \begin{pmatrix} 0 & -c_3 & c_2 \\ c_3 & 0 & -c_1 \\ -c_2 & c_1 & 0 \end{pmatrix}. \tag{28}$$

Using the Rodrigues' rotation formula [Rodrigues, 1840], we derive a closed form expression for the logarithmic map $\log_e : SO(3) \to \mathfrak{so}(3)$, with notation $\log_e g = (c_1, c_2, c_3)^\top$. We get[13]:

$$c_1 = \frac{\theta}{2\sin\theta}(\boldsymbol{G}_{g,3,2} - \boldsymbol{G}_{g,2,3}), \quad c_2 = \frac{\theta}{2\sin\theta}(\boldsymbol{G}_{g,2,1} - \boldsymbol{G}_{g,1,2}), \text{ and } \quad c_3 = \frac{\theta}{2\sin\theta}(\boldsymbol{G}_{g,1,3} - \boldsymbol{G}_{g,3,1}). \tag{29}$$

Due to the π periodicity of the orientation axis, the π-periodic Riemannian distance between two elements of the groups is the minimal distance between:

- the original distance without offset;
- the original distance with a negative offset $-\pi$ on the orientation axis;
- the original distance with a positive offset $+\pi$ on the orientation axis.

C Experiment details

Empirical evidence showed that deep and wide neural networks are keys to good performances. To help going deeper we use residual convolutional layers [He et al., 2016] and batch normalization [Ioffe and Szegedy, 2015]. To avoid tuning of the learning rate, we use ADAM optimizers [Kingma and Ba, 2014] for all our experiments. Last, we initialise our models' parameters via the Kaiming method [He et al., 2015].

Stability test on MNIST. In these experiments, we used a Wide Residual ChebNet with three convolutional residual layers with rectified linear units and kernel of size 4. As we do not pool at all, we only define one $SE(2)$ 16-NN graph with $28 \times 28 \times 6 = 4704$ vertices. We use extreme spatial anisotropy with $\epsilon^2 = 0.1$ and reach the 40/60 ratio by setting ξ accordingly. The residual layers are followed by a projective global max pooling layer and a fully connected with LogSoftMax output layer.

Orientation anisotropic test on CIFAR10. In these experiments, we used a Wide Residual ChebNet with three residual layers with rectified linear units and kernel of size 4. We use two R2RandPool layers. Hence we define three $SE(2)$ 16-NN graphs with $32 \times 32 \times 6 = 6144$, $16 \times 16 \times 6 = 1536$, and $8 \times 8 \times 6 = 384$ vertices. We use extreme spatial anisotropy with $\epsilon^2 = 0.1$. The residual layers are followed by a projective global max pooling layer and a fully connected with LogSoftMax output layer.

Spatial anisotropic test on CIFAR10. In these experiments, we used a Wide Residual ChebNet with 3 residual layers with rectified linear units and kernel of size 4. We use two R2RandPool layers. Hence we define three $SE(2)$ 16-NN graphs with $32 \times 32 \times 6 = 6144$, $16 \times 16 \times 6 = 1536$, and $8 \times 8 \times 6 = 384$ vertices. We use moderate orientation anisotropies to reach the 40/60 ratio on each graphs by setting ξ accordingly. The isotropic case is constructed using $\epsilon^2 = \xi^2 = 1$ on three $\mathbb{R}^2$

[12] Equiangular is far from uniform, but it has a sampling theorem.

[13] To compute the logarithmic map in the isotropic case corresponding to S^2, we must use a slightly modified version such that $c_3 = 0$, which via the exponential map generate torsion free exponential curves. The logarithmic of any rotation matrix defined by $(-\gamma, \beta, \gamma)$ yields this results [Bekkers, 2019, Portegies et al., 2015].

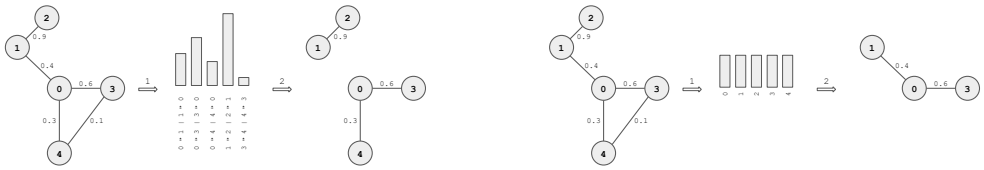
8-NN graphs with $32 \times 32 \times 1 = 784$, $16 \times 16 \times 1 = 256$, and $8 \times 8 \times 1 = 64$ vertices. The residual layers are followed by a projective global max pooling layer and a fully connected with LogSoftMax output layer.

Scalability test on STL10. In these experiments, we used a Wide Residual ChebNet with three residual layers with rectified linear units and kernel of size 4. We use two R2RandPool layers. Hence we define three $SE(2)$ 16-NN graphs with $96 \times 96 \times 6 = 55296$, $48 \times 48 \times 6 = 13824$, and $8 \times 8 \times 6 = 3456$ vertices. We use extreme spatial anisotropy with $\epsilon^2 = 0.1$ and reach the 40/60 ratio on each graphs by setting ξ accordingly. The isotropic case is constructed using $\epsilon = \xi = 1$ on three $\mathbb{R}^2$ 8-NN graphs with $96 \times 96 \times 1 = 9216$, $48 \times 48 \times 1 = 2304$, and $24 \times 24 \times 1 = 576$ vertices. The residual layers are followed by a projective global max pooling layer and a fully connected with LogSoftMax output layer. The residual layers are followed by a projective global max pooling layer and a fully connected with LogSoftMax output layer.

Scalability test on ClimateNet. In these experiments, we used a U-ChebNet with three residual layers with rectified linear units and kernel of size 3. We use five S2MaxPool (encoding) and five S2AvgUnpool (decoding) layers. Hence we define six $SO(3)$ 16-NN graphs with $10242 \times 6 = 61452$, $2562 \times 6 = 15372$, $642 \times 6 = 3852$, $162 \times 6 = 972$, $42 \times 6 = 252$, and $12 \times 6 = 72$ vertices. We use extreme spatial anisotropy with $\epsilon^2 = 0.1$ and reach the 40/60 ratio on each graphs by setting ξ accordingly. The residual layers are followed by a projective global max pooling layer and a fully connected with LogSoftMax output layer.

D Stability under random perturbations

In sensitive domains, the stability of a neural network to random perturbation is a desired property. We aim at demonstrate the high stability of our approach, as random perturbations are added to graphs during the training. We introduce two methods, both consisting in randomly pruning the original graph to construct a random sub-graph (see Figure 6). At test time, random perturbations are removed in order to evaluate a perturbation-free model.

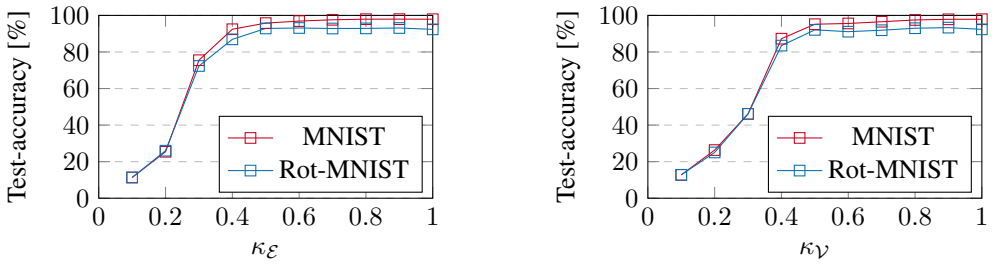


(a) One constructs a random sub-graph by randomly sampling a rate $\kappa_{\mathcal{E}}$ of edges of the complete graph based on their weight. As a direct consequence of the work of Keriven et al. [2020], the Laplacian of a quasi-sparse graph with randomly sampled edges remains consistent. Then, the unsampled edges are pruned by setting their weight to zero. One must be aware that this process could lead to isolated vertices or cluster of vertices. It is not necessarily a problem, but it is important to be mindful of this potential effect.

(b) This method based on vertices sampling consists of creating a random sub-graph by randomly sampling a rate $\kappa_{\mathcal{V}}$ of vertices from the complete graph. The edges between one or mode unsampled vertices are discarded too. This method is more drastic and should be used carefully, as such a vertices pruning leads to a violation of the uniform distribution of vertices.

Figure 6: Edge- and vertex-based sampling methods.

We run a bunch of experiments with a Wide Residual architecture on MNIST [LeCun and Cortes, 2010], varying the rates of edges or vertices to sample. The objective of these experiments is twofold. First, we would like to test the stability of the model under random perturbations. Second, this experiment is also a good manner to demonstrate the equivariance property of our method. In this purpose, we train our model on the original training set of MNIST, adding some perturbations. At test time, we evaluate the perturbation-free model on the test set with and without random rotations.



(a) Test-accuracy against rate of edges to sample

(b) Test-accuracy against rate of vertices to sample

Figure 7: Stability and group equivariance under random perturbations.

Equivariance. With these experiments, we empirically demonstrated the rotation-equivariance of a ChebLieNet. An extrinsic equivariance error can explain the slight difference in performance between the original test set and the randomly rotated one. This error is not related to the model but the evaluation's method and can be semantic (e.g. is it always possible to differentiate a six from a nine without orientation information) or numerical (e.g., how to rotate a low-resolution image by 10 degrees without alteration).

Robustness. While in this specific case, our dropout-like method does not help to improve our model on the test set, we expect it could help to reduce the over-fitting on other tasks. In addition, one could notice that even at middle-low sampling rates, the models remain stable. First of all, it supports the idea that there is no point using a fully connected graph; more edges capture the geometry better but with diminishing returns. Secondly, the model is robustly able to catch the semantic information in an image, even with partial alterations.

E Empirical convergence in eigenmaps of the graph Laplacians

As it has been proven by Belkin and Niyogi [2006], assuming a suitably constructed graph, the graph Laplacian converges in eigenmaps to its continuous counterpart, the Laplace-Beltrami operator. Hence, a good sanity check is to compare the eigenmaps of the discrete and continuous Laplace operators.

Firstly, we recall the eigenvalues of symmetric normalized graph Laplacians satisfies $0 \leq \lambda_k \leq 2$ for all k. As the eigenvalues can be interpreted as frequency components, the eigenvector associated to an high eigenvalue would correspond to an high frequency signal on the graph. To finish with, we propose a detailed analysis of the eigenmaps of our spaces of interest in terms of the Fourier basis and the spherical harmonics.

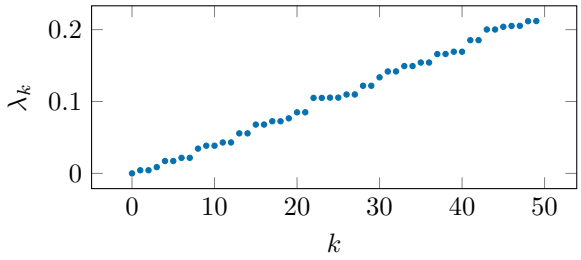
Isotropic 2-dimensional grid. The eigenmaps of the $[0, 1)^2$ space are shown in figure 8. It is a well-known fact that the 2-dimensional grid with periodic boundaries can be spanned by the Fourier basis. Consequently, the eigenvectors of this space corresponds to the sine and cosine trigonometric functions. In this case, the eigenvalues have multiplicity two (excepted the first one associated with a constant eigenvector), one for the sine and the other for the cosine. When the periodic condition at the boundaries are relaxed, the space is not homogeneous anymore, and the symmetries of the space change a little bit. The $[0, 1)^2$ space is such a space where periodic condition does not hold. Hence, the eigenmaps of this space are not equal to the Fourier basis, but very close.

Anisotropic 2-dimensional grid. The eigenmaps of the $[0, 1)^2 \times [-\pi/2, \pi/2)$ space are shown in figure 10. We can extend the discussion we add in the isotropic case to justify the fact that the eigenmaps of the $[0, 1)^2 \times [-\pi/2, \pi/2)$ is close to the Fourier basis. Moreover, we an additional orientation dimension, the rotation of the $[0, 1)^2$ induces a torsion all along the orientation axis because of the anisotropic Riemannian metric.

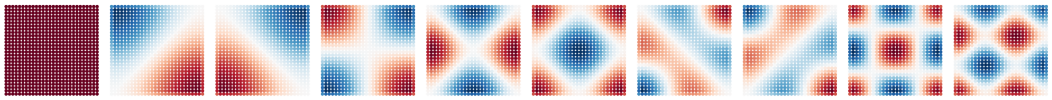
Isotropic 2-dimensional sphere. The eigenmaps of the S^2 space are shown in figure 9. It is a well-known fact that the 2-dimensional sphere is spanned by the spherical harmonics. The eigenvalues

associated to this eigenvectors have increasing multiplicities of the form $2m+1$ where m is the order of the spherical harmonic.

Anisotropic 2-dimensional sphere. The eigenmaps of the $S^2 \times [-\pi/2, \pi/2)$ space are shown in figure 11. We can extend the discussion we add in the isotropic case to justify the fact that the eigenmaps of the $S^2 \times [-\pi/2, \pi/2)$ are close to the spherical harmonics. Nevertheless, the role of the orientation dimension is not easily interpretable. Indeed the notion of orientation is arduous to understand, as the kernel's orientation depends on the path, and not on this orientation location.

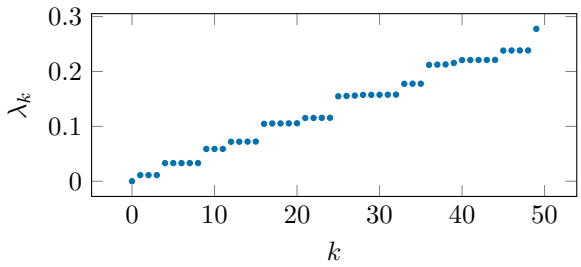


(a) Eigenvalues of the $[0,1]^2$ space, from λ_0 to λ_{49}.



(b) Eigenvectors of the $[0,1]^2$ space, from ϕ_0 (left) to ϕ_9 (right).

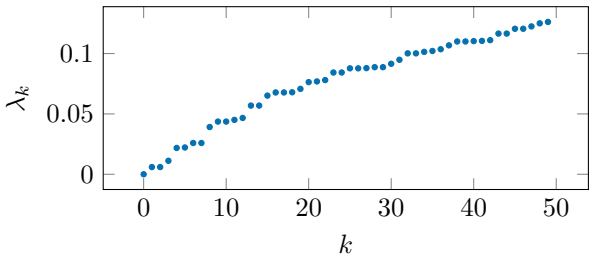
Figure 8: Eigenmaps of the $[0,1]^2$ space.



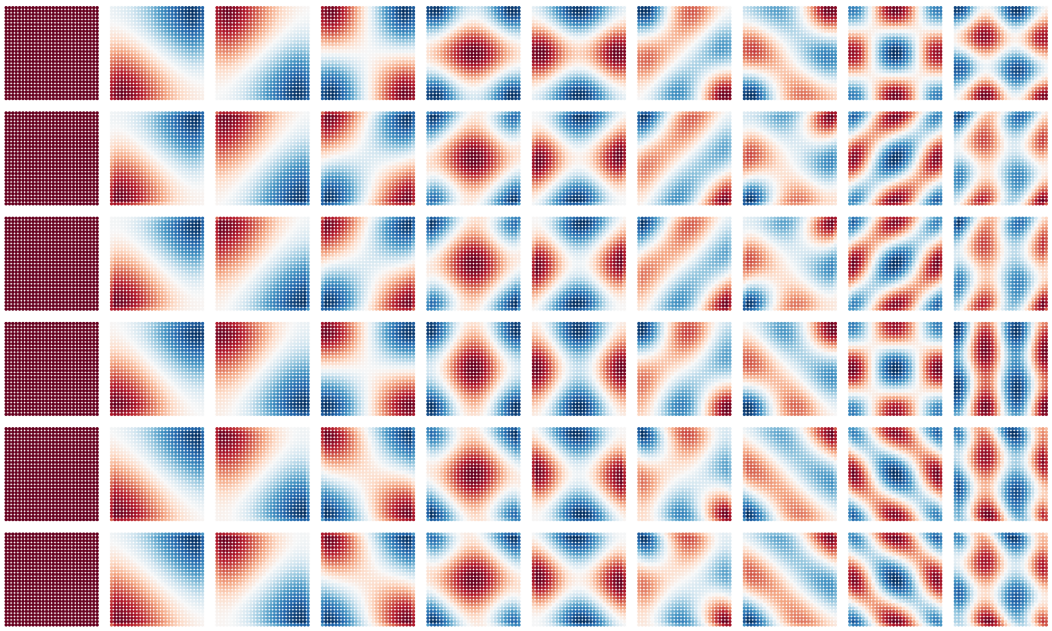
(a) Eigenvalues of the S^2 space, from λ_0 to λ_{49}.

(b) Eigenvectors of the S^2 space, from ϕ_0 (left) to ϕ_9 (right).

Figure 9: Eigenmaps of the S^2 space.

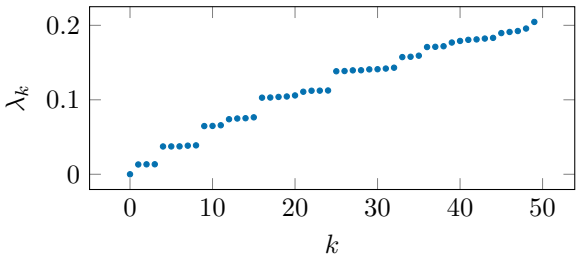


(a) Eigenvalues of the $[0,1]^2 \times [-\pi/2, \pi/2)$ space, from λ_0 to λ_{49}.

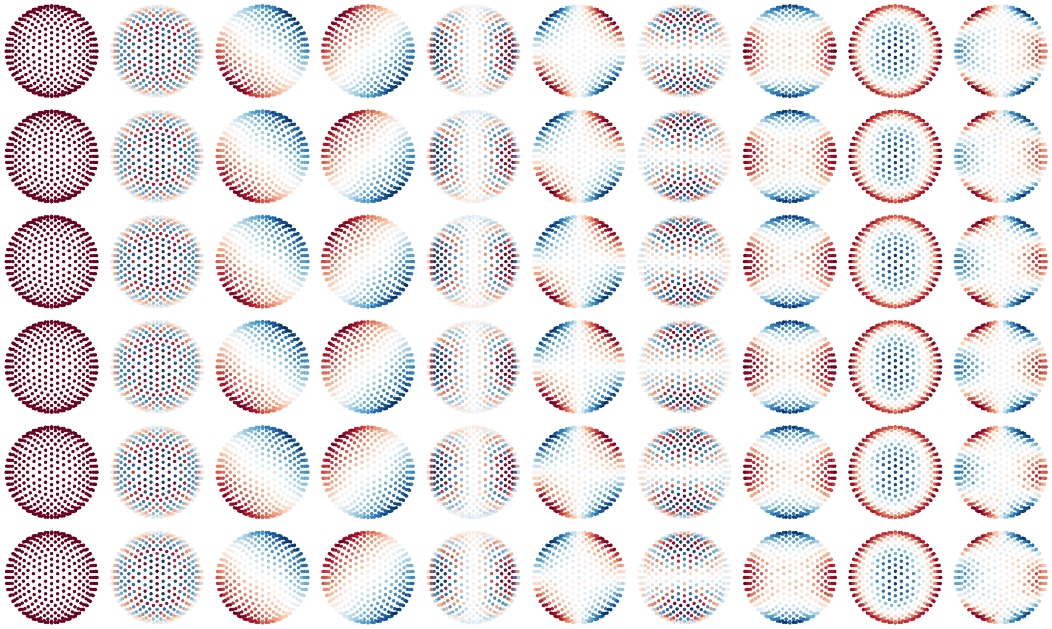


(b) Eigenvectors of the $[0,1]^2 \times [-\pi/2, \pi/2)$ space, from ϕ_0 (left) to ϕ_9 (right).

Figure 10: Eigenmaps of the $[0,1]^2 \times [-\pi/2, \pi/2)$ space.



(a) Eigenvalues of the $S^2 \times [-\pi/2, \pi/2)$ space, from λ_0 to λ_{49}.



(b) Eigenvectors of the $S^2 \times [-\pi/2, \pi/2)$ space, from ϕ_0 (left) to ϕ_9 (right).

Figure 11: Eigenmaps of the $S^2 \times [-\pi/2, \pi/2)$ space.