# Contents

Geometric Deep Learning
Beyond Euclidean Domains

# 1 Geometric Priors

**Definition 1.1.** *Our compact euclidean domain $\Omega$*
$\Omega := \prod_{i\in I}[0,1]$.

**Definition 1.2.** *Classification*
*Let $x \in L^2 := L^2(\Omega)$ then $f : L^2 \to \mathscr{C}$ surjective is said to be a classification of $L^2$ on the set $\mathscr{C}$.*

**Definition 1.3.** *Training set*
*Let $f$ be a classification of $L^2$ on $\mathscr{C}$ and $\{x_i\}_{i\in I} \subset L^2$ then the set $\{(x_i, f(x_i))\}_{i\in I}$ is called a training set for $f$.*

**Proposition 1.1.** *The classification $f$ is not injective*
*Let $f$ be a classification of $L^2$ on $\mathscr{C}$ then, given the inevitable noise acting on data, there exists a real positive $\varepsilon$ such that $\forall(x, x_\varepsilon) \in L^2 \times L^2 : \int\limits_\Omega |x - x_\varepsilon|^2 < \varepsilon$ we have that $f(x) = f(x_\varepsilon)$.*

Given ideal data classification we can define two functions $f$-equivalent if and only if their images via the classification $f$ are equal according to an equivalence on $\mathscr{C}$ which so far can be any set.

**Proposition 1.2.** *The relation $\simeq$ is an equivalence relation*
*Let $x, y, z \in L^2$ we define $x \simeq y \iff f(x) = f(y)$ where $f$ is a classification of $L^2$ on $\mathscr{C}$, then:*
*(i) $x \simeq x$*
*(ii) $x \simeq y \iff y \simeq x$*
*(iii) $x \simeq y, y \simeq z \implies x \simeq z$*

*Proof.* (i),(ii) and (iii) follow from the the equivalence on $\mathscr{C}$ by which they are defined. □

**Definition 1.4.** *Translation operator*
*Let $x \in L^2$ and $v \in \Omega$ then $T_v : L^2 \to L^2$ such that $x(\xi) \mapsto x(\xi - v)$ is said to be a translation operator.*

**Definition 1.5.** *Local deformation operator*
*Let $x \in L^2$ and $\tau \in C^\infty(\Omega, \Omega)$ then $L_\tau : L^2 \to L^2$ such that $x(\xi) \mapsto x(\xi - \tau(\xi))$ is said to be a local deformation operator according to the smooth vector field $\tau$.*

**Definition 1.6.** *Invariance*
*A classification $f$ of $L^2$ on $\mathscr{C}$ is said to be A-invariant, where $A : L^2 \to L^2$, if and only if $f(A(x)) = f(x)\ \forall x \in L^2$.*

**Definition 1.7.** *Equivariance*
*A classification $f$ of $L^2$ on $\mathscr{C}$ is said to be A-equivariant, where $A : L^2 \to L^2$, if and only if $f(A(x)) = A(f(x))\ \forall x \in L^2$. This is well defined only if A is defined to act on $\mathscr{C}$*

**Proposition 1.3.** *If $f$ is translation invariant then it is stable under local deformations*
*Let $f$ be a translation invariant classification of $L^2$ on $\mathscr{C}$ then $|f(L_\tau(x)) - f(x)| \approx |J_\tau|$ where $(J_\tau)_{ij} = (\frac{\partial \tau_i}{\partial \xi_j})$ under some misterious norm.*

*Proof.* To be found... □

## 2 Graphs and Manifolds

**Definition 2.1.** *Let $\mathcal{G}$ be a graph where $\mathcal{V}$ are its vertexes and $\mathcal{E}$ are its edges, let $f,g \in L^2(\mathcal{V})$ and $F,G \in L^2(\mathcal{E})$ be real valued functions, we define $\langle f,g\rangle_{L^2(\mathcal{V})} := \sum_{\mathcal{V}} a_i f_i g_i$, $a_i \in \mathbb{R}$ and $\langle F,G\rangle_{L^2(\mathcal{E})} := \sum_{\mathcal{E}} w_{ij} F_{ij} G_{ij}$, $w_{ij} \in \mathbb{R}$.*
*Let $M$ be a manifold and $TM$ its tangent bundle with a metric $\langle,\rangle_{TM} : TM^2 \to \mathbb{R}$, let $f,g \in L^2(M)$ and $F,G \in L^2(TM) := F : M \to TM$, given two scalar products $\langle f,g\rangle_{L^2(M)} := \int_M dx f g$ and $\langle F,G\rangle_{L^2(TM)} := \int_M dx \langle F,G\rangle_{TM}$.*

**Definition 2.2.** *Graph gradient and divergence*
*Let $f \in L^2(\mathcal{V})$ and $F \in L^2(\mathcal{E})$ we define $grad : L^2(\mathcal{V}) \to L^2(\mathcal{E})$ and $div : L^2(\mathcal{E}) \to L^2(\mathcal{V})$, such that $(grad f)_{ij} = f_i - f_j$ and $(div F)_i = \frac{1}{a_i} \sum_{j \in \mathcal{V}:(i,j)\in\mathcal{E}} w_{ij} F_{ij}$.*

**Proposition 2.1.** *Let $f \in L^2(\mathcal{V})$ and $F \in L^2(\mathcal{E}) : F_{ij} = -F_{ji}$ then $\langle f, div F\rangle_{L^2(\mathcal{V})} = \langle grad f, F\rangle_{L^2(\mathcal{E})}$, i.e. $div^\dagger = grad$.*

*Proof.* $\sum_{\mathcal{V}} a_i f_i (div F)_i = \sum_{\mathcal{E}} w_{ij} F_{ij} (f_i - f_j) = \sum_{i\in\mathcal{V}} \sum_{j\in\mathcal{V}:(i,j)\in\mathcal{E}} w_{ij} F_{ij} f_i$ thus $a_i (div F)_i = \sum_{j\in\mathcal{V}:(i,j)\in\mathcal{E}} w_{ij} F_{ij}$. □

**Definition 2.3.** *Manifold gradient and divergence*
*Let $f \in L^2(M)$ and $F \in L^2(TM)$ we define $(grad f)_i := \frac{\partial f}{\partial x_i}$ and $div F := \sum_i \frac{\partial F_i}{\partial x_i}$.*

**Proposition 2.2.** *Let $f \in L^2(M)$ and $F \in L^2(TM)$ then $\langle f, -div F\rangle_{L^2(M)} = \langle grad f, F\rangle_{L^2(TM)}$, i.e. $div^\dagger = grad$.*

*Proof.* $\int_M dx \langle grad f, F\rangle_{L^2(TM)} = \int_M dx (div(fF) - f div F) = \int_{\partial M} fF + \int_M dx f(-div F)$, where some condition must be added to make the first integral vanish. □

**Definition 2.4.** *Path on a graph*
*We define a curve parametrization on a graph be some injective function $\gamma : I \subset \mathbb{Z} \to \mathcal{V}$, we shall call $\gamma$ a curve meaning its image $\gamma(I)$, to satisfy the connectedness condition we want the simplicial complex $\{\gamma, [\gamma(n), \gamma(n+1)]_{n\in I}\}$ to have a 1 dimensional 0-homology group. We define path over a graph the sequence $\{[\gamma(n),\gamma(n+1)]\}_{n\in I} \subset \mathcal{E}$, we might as well call that $\gamma$ because we are bad people.*

**Theorem 2.3.** *Gradient theorem on graphs*
*Let $\gamma$ be a connected path on a graph, and $f \in L^2(\mathcal{V})$, than we have $\sum_\gamma (grad f) = Df_{\partial_0 \gamma}$, where we define $Df_{(i,j)} = f_i - f_j$, and by $\partial_0$ we mean the boundary operator.*

*Proof.* Left to the reader... □

**Theorem 2.4.** *Gauss theorem on graphs*
*Let $F \in L^2(\mathcal{E}) : F_{ij} = -F_{ji}$, let $\mathcal{A} \subset \mathcal{V}$ then if $a_i = w_{ij} = 1$ we have $\sum_{\mathcal{A}} (div F)_i = \sum_{\partial^0 \mathcal{A}} F_{ij}$.*

*Proof.* First of all we recall $\partial^0 \mathcal{A} = \{(i,j) \in \mathcal{E}, i \in \mathcal{A}, j \in \mathcal{V} \setminus \mathcal{A}\}$, then we see that $\sum_{\mathcal{A}} (div F)_i = \sum_{i\in\mathcal{A}} \sum_{j\in\mathcal{V}:(i,j)\in\mathcal{E}} F_{ij} = \sum_{i\in\mathcal{A}} \sum_{j\in\mathcal{V}\setminus\mathcal{A}:(i,j)\in\mathcal{E}} F_{ij} + \sum_{i\in\mathcal{A}} \sum_{j\in\mathcal{A}:(i,j)\in\mathcal{E}} F_{ij} = \sum_{\partial^0\mathcal{A}} F_{ij} + \sum_{(i,j)\in\mathcal{A}^2} adj(\mathcal{A})_{ij} F_{ij}$ where since $adj(\mathcal{A})_{ij} = adj(\mathcal{A})_{ji}$ we have by renaming dummy indexes $adj(\mathcal{A})_{ij} F_{ij} = -adj(\mathcal{A})_{ij} F_{ij} = 0$. □

**Proposition 2.5.** *The use of the coboundary operator makes sense only with antisimmetric functions on the edges, the antisimmetry of those function is somehow related to the orientation of surfaces.*

*Proof.* $\sum_{\partial^0 \sum_{i\in\mathcal{A}} i} F_{ij} = \sum_{\sum_{i\in\mathcal{A}} \partial^0 i} F_{ij}$ if an edge is in the coboundary of two different vertexes of $\mathcal{A}$ it will be count twice, that means zero times in $\mathbb{Z}_2$, similarly for that same edge we would sum $F_{ij} + F_{ji} = 0$. □

**Definition 2.5.** *Graph laplacian*
*Let $f \in L^2(\mathcal{V})$ we have that $\langle grad f, grad f\rangle = \langle div(grad f), f\rangle =: \langle \Delta f, f\rangle = \langle f, \Delta f\rangle$, where $\Delta : L^2(\mathcal{V}) \to L^2(\mathcal{V})$ is the Laplacian.*

**Proposition 2.6.** *The laplacian represents the difference between the function and a local average of the function*
*(i) Graphs*
*Let $w_{ii} = 0$, and $\sum_j w_{ij} = a_i$, i.e. normalized laplacian, we have that $(\Delta f)_i = f_i \frac{\sum_j w_{ij}}{a_i} - \sum_j \frac{w_{ij}}{a_i} f_j$ which is a weighted average.*
*(ii) Manifolds (Let's just see it on an euclidean domain)*
*Let $f_0 \int_{\partial B} dx - \int_{\partial B} dx f(x) \simeq \int_{\partial B} dx \langle grad f, x\rangle = \int_B dx \Delta f$, where $B$ is a ball centered in 0 (or at least has a boundary), we have that $f(0) - \frac{\int_{\partial B} dx f(x) f}{\int_{\partial B} dx} \simeq \frac{\int_B dx \Delta f}{\int_{\partial B} dx}$. Gauss theorem could be used since the incremental vector $x$ on a ball is parallel to $2x$ with is the gradient of the implicit function defining the ball.*

# 3 Spectrum of the laplacian

**Proposition 3.1.** *Variational problems related to the laplacian*
*Let $f \in L^2(M)$ the variational problem minimizing the Dirichlet energy functional $\int_M dx(gradf)^2$, admit as solution the kernel of the laplacian. Otherwise if we wish to have normalized functions only to avoid uniformly vanishing functions i.e. $\int_M dxf^2 = 1$ we get as solution the eigenfunction of the laplacian relative to its lowest eigenvalue. After this the same process can be applied to graphs.*

*Proof.* The proof is split in two different problems:

- $\int f\Delta f = \int (gradf)^2$, the problem $\delta \int (gradf)^2 = 0$ is solved with the Euler-Lagrange equations $div \frac{\partial \mathcal{L}}{\partial(gradf)} = \frac{\partial \mathcal{L}}{\partial f}$, which give $\Delta f = 0$.
- This time we have $\delta[\int (gradf)^2 - \lambda(\int f^2 - 1)] = 0$, which using the same equations leads to $\Delta f = \lambda f$, and since for those functions $\int f\Delta f = \lambda$, we want to find the minumum eigenvalue of the laplacian.

□

**Proposition 3.2.** *Discrete "variational" problems related to the laplacian*
*Let $\phi$ be an $k \times k$ matrix, the variational problem minimizing the Dirichlet energy $Tr(\phi^T \Delta\phi)$ is equivalent to the problem $\Delta\phi = 0$, while if we normalize the functions we get $\Delta\phi = \Lambda\phi$, where $\Lambda = diag(\lambda)$ with the k lowest eigenvalues.*

*Proof.* As follows:

- Let $min_\phi Tr(\phi^T \Delta\phi) = \phi_{li}\Delta_{lk}\phi_{ki} = \sum_i \phi_i^T \Delta\phi_i$ (Einstein notation) be our optimization problem, since $\Delta > 0$ we can simply solve for all $\phi_i$ and minimize $\frac{\partial \mathcal{L}_i}{\partial \phi_j} = \frac{\partial \phi_i^T \Delta\phi_i}{\partial \phi_j} = 2\Delta\phi_i = 0$.
- Let $min_\phi Tr(\phi^T \Delta\phi) = \phi_{li}\Delta_{lk}\phi_{ki} = \sum_i \phi_i^T \Delta\phi_i$ (Einstein notation) be our optimization problem, under the constraint $\phi^T\phi = I$(orthonormal) since $\Delta > 0$ we can simply solve for all $\phi_i$ and minimize $\frac{\partial \mathcal{L}_i}{\partial \phi_j} = \frac{\partial(\phi_i^T \Delta\phi_i - \lambda_i(\phi_i^T\phi_i - 1))}{\partial \phi_j}$(non sostituisco l'1 perché non posso sostituire il vincolo nel vincolo)$= 2\Delta\phi_i - 2\lambda_i\phi_i = 0$. And since the initial trace is equal to the sum of our k positive eigenvalues we shall take the k lowest eigenvalues. The problem can be shown to be $\Delta\phi = \Lambda\phi$.

□

**Proposition 3.3.** *Continuous spectrum of the laplacian in $\mathbb{R}^3$*
*Let $f \in L^2(\Omega)$, in particular in a Schwartz space where the Fourier Transform is invertible and selfadjoint, then we have that $\Delta\phi_p = p^2\phi_p$, where $\forall f \in L^2$ we can write $f = \int_\Omega dp c_p \phi_p$ and $\int_\Omega dx \phi_p \phi_{p'} = \delta(p - p')$, that is generalized Hilbert Base up to normalization.*

*Proof.* :
So if the Fourier transform is invertible we have that $f(x') = \int_\Omega dp e^{i\langle x',p\rangle} \int_\Omega dx e^{-i\langle x,p\rangle} f(x)$, then by the definition of the shift functional we can represent it with its integral kernel $\int_\Omega dp e^{i\langle x'-x,p\rangle}$, which is equal to $\delta(x - x') = \delta(x' - x)$. Trivially we have that our laplacian eigenfunctions are $\phi_p = e^{i\langle x,p\rangle}$ of eigenvalue $p^2$, since $f(x) = \mathcal{F}^\dagger \mathcal{F} f = \int_\Omega dp(\mathcal{F}f)(p)e^{i\langle x,p\rangle}$ we define our coefficients $c_p$. We just have to show that $\int dx e^{i\langle x,p'-p\rangle} = \delta(p' - p)$ that is true by definition.

□