# Differential Geometry for Image Processing Notes

31 October 2023

**Gijs Bellaard**
g.bellaard@tue.nl
Eindhoven University of Technology

## Contents

1. Introduction & Preliminaries ..... 1
1.1. Notation ..... 1
1.2. Linear Algebra ..... 1
1.3. Group Theory ..... 1
1.4. Calculus ..... 2
1.5. Vector Calculus ..... 2
1.6. Functional Analysis ..... 3
1.7. Function Spaces ..... 5
1.8. Fourier Transform ..... 5
1.9. Differential Geometry ..... 6
2. Geometric Image Processing ..... 8
2.1. Gaussian Scale Space ..... 8
2.2. Poisson Scale Space ..... 8
2.3. Gaussian Scale Space on Bounded Domain ..... 10
2.4. Gaussian Derivatives ..... 11
2.5. Hermite Polynomials ..... 12
2.6. Separable and Isotropic Filters ..... 12
2.7. B-Splines ..... 14
2.8. Linear Least Squares ..... 15
2.9. Operator Exponential, One Parameter Semi-Groups, and Generators ..... 16
2.10. Roto-Translations ..... 17
2.11. Lie Group Theory ..... 19
2.12. Linear Constrained Optimization & Lagrange Multipliers ..... 21
2.13. Variational Techniques for Image Denoising ..... 21
3. Invertible Orientation Scores ..... 22
3.1. Group Representation ..... 22
3.2. Wavelet Transform ..... 23
3.3. Integer Shift Transform ..... 25
3.4. Roto-Translation Transform ..... 26
3.5. Similarity Transform ..... 26
3.6. Orientation Score Transform ..... 28
3.7. Reproducing Kernel Hilbert Spaces ..... 29
3.8. Space of Bandlimited Square Integrable Functions . . 31
3.9. Wavelet Transform on RKHS's 🚧 ..... 32
4. Tracking in Orientation Scores ..... 32
4.1. Finsler Geometry ..... 32
4.2. Reeds-Shepp Car Model ..... 33
Bibliography ..... 34

# 1. Introduction & Preliminaries

These are notes for Remco Duits' master course *"Differential Geometry for Image Processing"*. These notes do not strictly follow the structure of the official lecture notes. It contains most of the exercises and their answers, but almost always spread over the place.

## 1.1. Notation

Notation is often shortened when no confusion can arise, for example

- $\frac{\partial}{\partial x} = \partial_x$.
- $\frac{\partial}{\partial t} f(t) = \dot{f}(t)$ for time derivatives we sometimes use *Newton dot notation.*
- $\exp(x) = e^x$ even for generalized exponentials.
- $A(x) = Ax$ for linear maps $A$.
- $A \circ B = AB$ for linear maps $A, B$.
- $\int_\Omega f(x)\mu(dx) = \int_\Omega f \, \mathrm{d}\mu$ if the variable over which we integrate is clear from context.
- $\int_\Omega f(x)\mu(dx) = \int_\Omega f(x) \, \mathrm{d}x$ if the measure used is clear from context.
- $\int_\Omega f(x)\mu(dx) = \int f(x)\mu(dx)$ if the domain $\Omega$ that is used is clear from context.
- $\langle \cdot, \cdot \rangle_V = \langle \cdot, \cdot \rangle$ if the inner product space $V$ that is used is clear from context.
- $\hat{f} = \mathcal{F} f$ where $\mathcal{F}$ is a Fourier transform.
- $\mathbb{F}$ will indicate a generic field, usually $\mathbb{R}$ or $\mathbb{C}$.

## 1.2. Linear Algebra

**Definition 1.2.1 (2D Rotation Matrix)**: Let $R_\theta \in \mathbb{R}^{2\times 2}$ be the two dimensional rotation matrix which rotates anticlockwise with $\theta \in \mathbb{R}$, and $X \in \mathbb{R}^{2\times 2}$ as follows:

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \text{ and } X = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

## 1.3. Group Theory

**Definition 1.3.1 (Left/Right-Translation)**: Let $G$ be a group. We define the *left-translation* $L_g : G \to G$ by

$$L_g(h) = gh,$$

and the *right-translation* $R_g : G \to G$ by

$$R_g(h) = hg^{-1}$$

**Lemma 1.3.1**: The left and right translations satisfy

$$L_{g_1} \circ L_{g_2} = L_{g_1 g_2}, \quad R_{g_1} \circ R_{g_2} = R_{g_1 g_2}$$

**Definition 1.3.2 (Left/Right Translation of Functions)**: Let $G$ be a group and $\mathbb{F}$ any field. We can define a natural left-translation $\mathcal{L}_g$ of scalar functions $f : G \to \mathbb{F}$ by defining:

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

Similarly, we can define a natural right-translation $\mathcal{R}_g$ of functions $f : G \to \mathbb{F}$ by defining:

$$(\mathcal{R}_g f)(h) = f(hg)$$

Equivalently we can write

$$\mathcal{L}_g f = f \circ L_{g^{-1}}$$
$$\mathcal{R}_g f = f \circ R_{g^{-1}}$$

The mappings $\mathcal{L}$ and $\mathcal{R}$ are also called the *left- and right-regular representation* of $G$.

**Lemma 1.3.2**: The left and right translation operators $\mathcal{L}$ and $\mathcal{R}$ satisfy

$$\mathcal{L}_{g_1} \circ \mathcal{L}_{g_2} = \mathcal{L}_{g_1 g_2}, \quad \mathcal{R}_{g_1} \circ \mathcal{R}_{g_2} = \mathcal{R}_{g_1 g_2}$$

*Proof*: Let $f \in G \to \mathbb{F}$ be some dummy function. We will straightforwardly use properties of left and right translation $L_g$ and $R_g$.

$$\begin{aligned}\left(\mathcal{L}_{g_1}\left(\mathcal{L}_{g_2} f\right)\right) &= \left(\mathcal{L}_{g_2} f\right) \circ L_{g_1^{-1}} = f \circ L_{g_2^{-1}} \circ L_{g_1^{-1}} \\ &= f \circ L_{g_2^{-1} g_1^{-1}} = f \circ L_{(g_1 g_2)^{-1}} \\ &= \left(\mathcal{L}_{g_1 g_2} f\right)(h)\end{aligned}$$

The proof for $\mathcal{R}$ goes exactly the same way. □

## 1.4. Calculus

**Definition 1.4.1 (Taylor Series)**:

$$f(x+h) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x)}{n!} h^n$$

**Lemma 1.4.1 (Some Taylor Series)**:

$$\cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 + \cdots$$

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 + \cdots$$

**Definition 1.4.2 (Polar Coordinates)**:

$$x = r\cos\theta \quad \text{and} \quad y = r\sin\theta$$

$$r = \sqrt{x^2 + y^2} \quad \text{and} \quad \theta = \arctan\left(\frac{y}{x}\right)$$

**Proposition 1.4.1 (Polar Jacobians)**:

$$\begin{aligned}\frac{\partial(x,y)}{\partial(r,\theta)} &= \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} \\ &= \begin{pmatrix} \frac{x}{r} & -y \\ \frac{y}{r} & x \end{pmatrix}\end{aligned}$$

$$\begin{aligned}\frac{\partial(r,\theta)}{\partial(x,y)} &= \begin{pmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \end{pmatrix} = \begin{pmatrix} \frac{x}{r} & \frac{y}{r} \\ -\frac{y}{r^2} & \frac{x}{r^2} \end{pmatrix} \\ &= \begin{pmatrix} \cos\theta & -\sin\theta \\ -\frac{1}{r}\sin\theta & \frac{1}{r}\cos\theta \end{pmatrix}\end{aligned}$$

**Proposition 1.4.2 (Polar Differentials)**:

$$\partial_r = +\frac{x}{r}\partial_x + \frac{y}{r}\partial_y$$
$$\partial_\theta = -y\,\partial_x + x\,\partial_y$$

$$\partial_x = +\cos\theta\partial_r - \frac{1}{r}\sin\theta\,\partial_\theta$$
$$\partial_y = -\sin\theta\,\partial_r + \frac{1}{r}\cos\theta\partial_\theta$$

## 1.5. Vector Calculus

**Definition 1.5.1 (Gradient)**: The *gradient* $\nabla f$ of a differentiable scalar field $f \in C^1\left(\mathbb{R}^d \to \mathbb{R}\right)$ is

$$(\nabla f)(x) = \begin{pmatrix} \partial_1 f \\ \partial_2 f \\ \vdots \\ \partial_d f \end{pmatrix}$$

**Definition 1.5.2 (Curl)**: In three dimensions, the *curl* $\nabla \times F$ of a differentiable vector field $F \in C^1\left(\mathbb{R}^d \to \mathbb{R}^d\right)$ is.

$$(\nabla \times f)(x) = \begin{pmatrix} \partial_2 F_3 - \partial_3 F_2 \\ \partial_3 F_1 - \partial_1 F_3 \\ \partial_1 F_2 - \partial_2 F_1 \end{pmatrix}$$

**Definition 1.5.3 (Divergence)**: The *divergence* $\nabla \cdot F$ of a differentiable vector field $F \in C^1\left(\mathbb{R}^d \to \mathbb{R}^d\right)$ is

$$\nabla \cdot F = \nabla \cdot \nabla F = \partial_1 F_1 + \partial_2 F_2 + \cdots + \partial_d F_d$$

**Definition 1.5.4 (Laplacian)**: The *Laplacian* $\Delta f$ of a twice differentiable scalar field $f : C^2\left(\mathbb{R}^d \to \mathbb{R}\right)$ is

$$\Delta f = (\partial_1)^2 f + (\partial_2)^2 f + \cdots + (\partial_d)^2 f. \tag{1}$$

**Lemma 1.5.1 (Divergence Theorem)**: Let $\Omega \subset \mathbb{R}^d$ be a bounded open set with its boundary $\partial\Omega$ being $C^1$. Let $F \in C^1\left(\overline{\Omega} \to \mathbb{R}^d\right)$ be a vector field. The *divergence theorem* states that

$$\int_\Omega \nabla \cdot F = \int_{\partial\Omega} F \cdot n$$

where $n : \partial\Omega \to \mathbb{R}^d$ is the outward pointing unit normal at each point on the boundary.

**Lemma 1.5.2 (Vector Product Rule)**:

$$\nabla \cdot (f \nabla g) = (\nabla f) \cdot (\nabla g) + f \Delta g$$

**Lemma 1.5.3 (Greens First Identity)**: Let $f, g \in C^2\left(\mathbb{R}^d \to \mathbb{R}\right)$.

$$\int_\Omega (\nabla f) \cdot (\nabla g) + f \Delta g = \int_\Omega \nabla \cdot (f \nabla g) = \int_{\partial\Omega} (f \nabla g) \cdot n \tag{2}$$

## 1.6. Functional Analysis

**Definition 1.6.1 (Metric & Metric Space)**: Let $X$ be a set. A *metric* on $X$ is a mapping $d(\cdot, \cdot) : X \times X \to \mathbb{R}_{\geq 0}$ satisfying

1. Positive Definiteness: $d(x, y) = 0 \Leftrightarrow x = y$
1. Symmetry: $d(x, y) = d(y, x)$
2. Triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$

A *metric space* is a set with a metric.

**Definition 1.6.2 (Continuity on Metric Spaces)**: Let $X, Y$ be two metric spaces. Let $f : X \to Y$ be a function. The function is called *continuous* at $x \in X$ if for every $\varepsilon > 0$ we can find a $\delta > 0$ such that for all $d(x', x) < \delta$ one has that $d(f(x'), f(x)) < \varepsilon$. The function is called continuous if it is continuous everywhere.

**Definition 1.6.3 (Distance Preserving)**: Let $X, Y$ be two metric spaces. A mapping $f : X \to Y$ is called *distance preserving*[1] if

$$d(x_1, x_2) = d(f(x_1), f(x_2))$$

**Definition 1.6.4 (Norm & Normed Space)**: Let $X$ be a (possibly infinite dimensional) vector space. A *norm* on $X$ is a mapping $\|\cdot\| : X \to \mathbb{R}_{\geq 0}$ satisfying

1. Positive Definiteness: $x \neq 0 \Rightarrow \|x\| > 0$
2. Absolute Homogeneity: $\|\lambda x\| = |\lambda| \|x\|$
3. Triangle inequality: $\|a + b\| \leq \|a\| + \|b\|$

A *normed space* is a vector space with a norm.

Every normed space is a metric space by defining $d(x, y) = \|x - y\|$.

**Definition 1.6.5 (Bounded Linear Operator)**: Let $X, Y$ be two normed spaces. Let $A : X \to Y$ be a linear map. Then $A$ is *bounded* if their exists some $M \geq 0$ such that

$$\|Ax\| \leq M \|x\|$$

The set of all bounded linear operators between $X$ and $Y$ is denoted with $B(X, Y)$, and the set of all bounded linear operators between $X$ and itself is denoted with $B(X)$.

**Definition 1.6.6 (Norm Preserving)**: Let $X, Y$ be two normed spaces. Let $A : X \to Y$ be a linear map. Then $A$ is *norm preserving*[2] if

$$\|Ax\| = \|x\|$$

for all $x \in X$.

**Lemma 1.6.1 (Boundedness and Continuity)**: Let $X, Y$ be two normed spaces. Let $A : X \to Y$ be a linear map. Then $A$ is continuous if and only if it is bounded.

*Proof*: Let us first show that boundedness implies continuity. So, assume that $A$ is bounded. We want to show that $A$ is continuous. This means that for all $x \in X$ and all $\varepsilon > 0$ we can find a $\delta > 0$ such that for all $x' \in X$

$$\|x' - x\| < \delta \Rightarrow \|Ax' - Ax\| < \varepsilon.$$

Let us inspect $\|Ax' - Ax\|$

$$\|Ax' - Ax\| = \|A(x' - x)\| \leq M \|x' - x\|$$

so we can choose $\delta = \frac{\varepsilon}{M}$, giving

[1] This is also called an *isometry*.
[2] This is also called a *linear isometry*.

$$\|Ax' - Ax\| \leq M\|x' - x\| < M\delta = M\frac{\varepsilon}{M} = \varepsilon$$

as desired.

Now let us now show that continuity implies boundedness. So, assume that $A$ is continuous. We want to show that $A$ is bounded. $A$ is continuous so it is continuous at $0$. Let us pick $\varepsilon = 1$, we know that their exists some $\delta > 0$ such that

$$\|x\| < \delta \Rightarrow \|Ax\| < \varepsilon = 1$$

Now let us inspect $\|Ax\|$ for a general $x \in X$.

$$\begin{aligned}\|Ax\| &= \left\|A\frac{\delta x}{\|x\|}\frac{\|x\|}{\delta}\right\| = \frac{\|x\|}{\delta}\left\|A\frac{\delta x}{\|x\|}\right\| \\ &< \frac{\|x\|}{\delta}\varepsilon = \frac{\|x\|}{\delta}\end{aligned}$$

We see that $M = \frac{1}{\delta}$ can be used to bound $A$, and thus $A$ is bounded.

□

**Definition 1.6.7 (Cauchy Sequence & Complete Space)**: Let $(X, d)$ be a metric space. A sequence $x_1, x_2, \ldots \in X$ is called *Cauchy* if for every $\varepsilon > 0$ there exists a $N$ such that for all $n, m > N$ one has $d(x_m, x_n) < \varepsilon$.

A metric space is called *complete* (or *Cauchy*) if every Cauchy sequence has a limit in that space.

**Definition 1.6.8 (Banach Space)**: A *Banach space* is a complete normed space.

**Definition 1.6.9 (Absolute Convergence)**: Let $(X, \|\cdot\|)$ be a normed space. A series $\sum x_i$ is called *absolutely convergent* if the series $\sum\|x_i\|$ converges in $\mathbb{R}$.

**Lemma 1.6.2 (Absolute Convergence implies Convergence)**: Let $(X, \|\cdot\|)$ be a Banach space. Let $\sum x_i$ be an absolutely convergent series. Then $\sum x_i$ converges.

*Proof*: By assumption $\sum\|x_i\|$ converges. This means that the sequence of partial sums $\sum_{i=1}^{n}\|x_i\|$ is Cauchy. By definition of a Cauchy sequence, this means that for every $\varepsilon > 0$ one can find a $N$ such that for al $m > n > N$ one has

$$\sum_{i=1}^{m}\|x_i\| - \sum_{i=1}^{n}\|x_i\| < \varepsilon.$$

We will now show using this that the sequence of partial sums $\sum_{i=1}^{n} x_i$ is Cauchy.

$$\begin{aligned}\left\|\sum_{i=1}^{m} x_i - \sum_{i=1}^{n} x_i\right\| &= \left\|\sum_{i=n+1}^{m} x_i\right\| \leq \sum_{i=n+1}^{m}\|x_i\| \\ &= \sum_{i=1}^{m}\|x_i\| - \sum_{i=1}^{n}\|x_i\| < \varepsilon\end{aligned}$$

So the sequence of partial sums $\sum_{i=1}^{n} x_i$ is Cauchy and so, by definition of a Banach space, this converges within the space.

□

**Definition 1.6.10 (Inner Product & Inner Product Space)**: Let $X$ be a (possibly infinite dimensional) vector space over a field $\mathbb{F}$ that is either $\mathbb{R}$ or $\mathbb{C}$. An *inner product* on $X$ is a mapping $\langle\cdot,\cdot\rangle : X \times X \to \mathbb{F}$ satisfying

1. Conjugate Symmetry: $\langle x, y\rangle = \overline{\langle y, x\rangle}$
2. Positive Definiteness: $x \neq 0 \Rightarrow \langle x, x\rangle > 0$
3. Linear: $\langle x, ay + bz\rangle = a\langle x, y\rangle + b\langle x, z\rangle$
4. Conjugate-Linear: $\langle ax + by, z\rangle = \bar{a}\langle x, z\rangle + \bar{b}\langle y, z\rangle$

A *inner product space* is a vector space with an inner product.

Note that we use the convention that the inner product is linear in the second argument.

Every inner product space is a normed space by defining $\|x\| = \sqrt{\langle x, x\rangle}$. This is in fact not immediately clear: one needs the *Cauchy–Schwarz inequality* to proof that this norm satisfies the triangle inequality.

**Lemma 1.6.3 (Cauchy-Schwartz Inequality)**: Let $X$ be a inner product space. One has the *Cauchy–Schwarz inequality*:

$$|\langle x, y\rangle| \leq \|x\|\|y\|$$

**Definition 1.6.11 (Inner Product Preserving)**: Let $X, Y$ be two inner product spaces. A linear map $A : X \to Y$ is called *inner product preserving* if

$$\langle x_1, x_2\rangle_X = \langle Ax_1, Ax_2\rangle_Y$$

It is immediately clear that an inner product preserving map also preserves the the associated norms. However, in fact, a norm preserving map between inner product spaces is also inner product preserving.

**Lemma 1.6.4 (Norm Preserving is Inner Product Preserving)**: Let $X, Y$ be two inner product spaces with their associated norms. Let $A : X \to Y$ be a linear norm preserving map. Then $A$ is also inner product preserving.

*Proof*: This follows directly from the identity

$$4\langle x, y\rangle = \|x + y\|^2 - \|x - y\|^2$$

in the real case, and the identity

$$4\langle x, y\rangle = \|x+y\|^2 - \|x-y\|^2 - i\|x+iy\|^2 + i\|x-iy\|^2$$

in the complex case. Both can be checked using $\|\cdot\|^2 = \langle \cdot, \cdot \rangle$. The identity is useful because it relates the inner product purely in terms of norms. Because $A$ is norm preserving we now see that it is also inner product preserving. □

**Definition 1.6.12 (Hilbert Space)**: A *Hilbert space* is a complete inner product space. Every Hilbert space is Banach.

**Lemma 1.6.5 (Unitary Operator)**: Let $H$ be a Hilbert space. A *unitary operator* $U : H \to H$ is a mapping that is bijective and preserves the inner product.

**Definition 1.6.13 (Adjoint Operator)**: Let $A : H_1 \to H_2$ be a bounded linear operator between two Hilbert spaces $H_1, H_2$. There exists a bounded operator $A^* : H_2 \to H_1$, called the *adjoint* of $A$ such that for all $x \in H_1$ and $y \in H_2$ one has

$$\langle Ax, y\rangle_{H_2} = \langle x, A^* y\rangle_{H_1},$$

The existence and uniqueness of the adjoint follows from the Riesz representation theorem. If $A$ is a bounded linear operator, the adjoint is also bounded.

## 1.7. Function Spaces

**Definition 1.7.1 (Continuous Differentiable Space)**: A function $f : \mathbb{R} \to \mathbb{C}$ is said to be $k$-times continuously differentiable if the derivatives $f, f', \ldots, f^{(k)}$ exist and are continuous. It is said to be *smooth* if *all* derivatives $f, f', f'', \ldots$ exist and are continuous.

The space $C^k(\mathbb{R})$ is defined as the vector space of all $k$-times continuously differentiable functions. The space $C^\infty(\mathbb{R}^d)$ is defined as the vector space of all smooth functions.

We generalize these notions straightforwardly to functions $f : \mathbb{R}^d \to \mathbb{C}$.

**Definition 1.7.2 (p (Semi-)Norm)**: Let $(X, \mu)$ be a measure space. We define the *p-(semi-)norm* $\|f\|_p$ of a measurable function $f : X \to \mathbb{F}$ as

$$\|f\|_p = \sqrt[p]{\int |f|^p \, \mathrm{d}\mu}$$

where $\int$ is the *Lebesque* integral, and we allow the (semi-)norm to be infinite.

**Definition 1.7.3 (Lebesque Space)**: Let $(X, \mu)$ be a measure space, $M(X, \mathbb{F})$ the space of measurable functions $f : X \to \mathbb{F}$, and $p \geq 1$. We define the space $\mathcal{L}^p(X)$ as

$$\mathcal{L}^p(X) = \left\{ f \in M(X, \mathbb{F}) \mid \|f\|_p < \infty \right\}$$

This space might in fact be a *semi*-normed space as the Lebesque integral does not care what happens on sets of measure zero. To turn it into a normed space for certain we qoutient out the null set

$$\mathcal{N}^p(X) = \left\{ f \in M(X, \mathbb{F}) \mid \|f\|_p = 0 \right\}$$

After we qoutient out the problem-makers we obtain *Lebesque p space* $\mathbb{L}^p(X)$:

$$\mathbb{L}^p(X) = \mathcal{L}^p(X)/\mathcal{N}^p(X)$$

which is a proper normed vector space.

All Lebesque spaces are Banach spaces, and $\mathbb{L}^2$ space is even a Hilbert space.

**Definition 1.7.4 (Schwartz Space)**: *Schwartz space* $S(\mathbb{R}^d)$ is defined as

$$S(\mathbb{R}^d) = \left\{ f \in C^\infty(\mathbb{R}^d) \mid \forall \alpha, \beta : \|f\|_{\alpha,\beta} < \infty \right\}$$

where $a, b \in \mathbb{N}^d$ are multi-indices, and

$$\|f\|_{\alpha,\beta} := \sup_{x \in \mathbb{R}^d} |x^a (D^\beta f)(x)|$$

Note that Schwartz is **not** a Banach space because it is not *normable*. Schwartz space is an **extremely** nice space of functions. The functions are infinity differentiable, i.e. *smooth*, and every possible derivative decays faster than any polynomial, i.e. *rapidly decreasing*. Schwartz space is a natural space to use when working with PDE's and Fourier transformation. Namely, the space is *closed* under differentiation, multiplication with polynomials, multplication of functions, convolutions of functions, and taking the Fourier transform. Schwartz space is usefull because it also dense in the Lebesque spaces $\mathbb{L}^p(\mathbb{R}^d)$ space for all $p \geq 1$.

## 1.8. Fourier Transform

Here we introduce the Fourier transform on Schwartz space because it makes our lives way easier. One doesn't need to consider such an extremely nice space neccesarily. For example, the Fourier transform can also be introduced as a transform taking $\mathbb{L}^1(\mathbb{R}^d)$ into $C_0(\mathbb{R}^d)$. Note however that, to my knowledge, this mapping is **not** onto, just into, and there is **no** characterization of the image in this case. There is no "best" class of functions to consider so several variants for the domain/codomain of the Fourier transform exist.

**Definition 1.8.1 (Fourier Transform)**: The *Fourier transform* $\mathcal{F} : S(\mathbb{R}^d) \to S(\mathbb{R}^d)$ is an bijective linear operator defined by

$$\hat{f}(\omega) = (\mathcal{F}f)(\omega) = \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} f(x) e^{-i\omega \cdot x} \, \mathrm{d}x. \tag{3}$$

The inverse transform $\mathcal{F}^{-1} : S(\mathbb{R}^d) \to S(\mathbb{R}^d)$ is

$$f(x) = \left(\mathcal{F}^{-1}\hat{f}\right)(x) = \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} \hat{f}(\omega) e^{i\omega \cdot x} \, \mathrm{d}\omega. \tag{4}$$

Proving that the Fourier transform in Schwarz space is indeed bijective, and that the inverse is as above, is not trivial. Must proofs of the *Fourier inversion theorem* you find online will give a fake proof using the Dirac delta "function". I refer to Frederick Schuller's video lecture[3] for a very clear and mathematicaly precise introduction to the Fourier operator.

An useful property of the Fourier transform is the *Convolution theorem*:

**Lemma 1.8.1 (Convolution Theorem)**: Let $f, g \in S(\mathbb{R}^d)$ then

$$\mathcal{F}(f * g) = (2\pi)^{\frac{d}{2}} (\mathcal{F}f) \cdot (\mathcal{F}g),$$

That is, the Fourier transform turns convolution $*$ in the spatial domain into multiplication $\cdot$ in the frequency domain. Equivalently:

$$\mathcal{F}^{-1}(f \cdot g) = \frac{1}{(2\pi)^{\frac{d}{2}}} (\mathcal{F}^{-1}f) * (\mathcal{F}^{-1}g).$$

Or in short

$$f * g \longleftrightarrow (2\pi)^{\frac{d}{2}} \hat{f} \cdot \hat{g}$$

The Fourier transform also turns spatial derivatives into multiplications.

**Lemma 1.8.2 (Derivatives and Multiplications)**: Let $f \in S(\mathbb{R}^d)$ then

$$\mathcal{F}(\partial_j f)(\omega) = (i\omega_j) \cdot (\mathcal{F}f)(\omega) \tag{5}$$

*Proof*: We will only show this for the one dimensional case $d = 1$. Consider

$$f(x) = \left(\mathcal{F}^{-1}\hat{f}\right)(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \hat{f}(\omega) e^{i\omega x} \, \mathrm{d}\omega$$

Taking the derivative on both sides gives

$$(\partial f)(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} i\omega \hat{f}(\omega) e^{i\omega x} \, \mathrm{d}\omega = \mathcal{F}^{-1}\left(i\omega \hat{f}\right)$$

Taking the Fourier transform on both sides gives the result. □

**Lemma 1.8.3 (Plancherel theorem)**: The Fourier transform $\mathcal{F} : S(\mathbb{R}^d) \to S(\mathbb{R}^d)$ is $\mathbb{L}^2(\mathbb{R}^d)$ inner product preserving

$$\langle \mathcal{F}f_1, \mathcal{F}f_2 \rangle_2 = \langle f_1, f_2 \rangle_2$$

*Proof*: Let $f_1, f_2 \in S(\mathbb{R}^d)$. We already know that $f_1, f_2 \in (\mathbb{L}^1 \cap \mathbb{L}^2)(\mathbb{R}^d)$ due to Schwarz space being a subset of both. This makes all upcoming manipulations justified. Most references derive the equality using the Dirac delta "function" which in our context is completely unjustified. We will abuse that we already know that $\mathcal{F}^{-1}\mathcal{F}f = f$

$$\begin{aligned}
&\langle \mathcal{F}f_1, \mathcal{F}f_2 \rangle_2 \\
&= \int_{\mathbb{R}^d} \overline{(\mathcal{F}f_1)(\omega)} (\mathcal{F}f_2)(\omega) \, \mathrm{d}\omega \\
&= \int_{\mathbb{R}^d} \overline{\int_{\mathbb{R}^d} f_1(x) e^{-i\omega \cdot x} \, \mathrm{d}x} (\mathcal{F}f_2)(\omega) \, \mathrm{d}\omega \\
&= \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} \overline{f_1(x)} e^{i\omega \cdot x} \, \mathrm{d}x \right) (\mathcal{F}f_2)(\omega) \, \mathrm{d}\omega \\
&= \int_{\mathbb{R}^d} \overline{f_1(x)} \int_{\mathbb{R}^d} (\mathcal{F}f_2)(\omega) e^{i\omega \cdot x} \, \mathrm{d}\omega \, \mathrm{d}x \\
&= \int_{\mathbb{R}^d} \overline{f_1(x)} (\mathcal{F}^{-1}\mathcal{F}f_2)(x) \, \mathrm{d}x \\
&= \int_{\mathbb{R}^d} \overline{f_1(x)} f_2(x) \, \mathrm{d}x \\
&= \langle f_1, f_2 \rangle_2
\end{aligned}$$

□

**Definition 1.8.2 (Extended Unitary Fourier)**: Now that we know that $\mathcal{F} : S(\mathbb{R}^d) \to S(\mathbb{R}^d)$ is a densely defined inner product preserving operator from $\mathbb{L}^2(\mathbb{R}^d)$ to itself, we know of the existence of extension $\mathcal{F} : \mathbb{L}^2(\mathbb{R}^d) \to \mathbb{L}^2(\mathbb{R}^d)$ that is *unitary*.

In $\mathbb{L}^2(\mathbb{R}^d)$ the formula for the Fourier transform in $S(\mathbb{R}^d)$ is technically not correct anymore, but the following is:

$$\mathcal{F}f = \frac{1}{(2\pi)^{\frac{d}{2}}} \lim_{R \to \infty} \left( \omega \mapsto \int_{B_R} f(x) e^{-i\omega \cdot x} \, \mathrm{d}x \right)$$

where the limit is in the $\mathbb{L}^2(\mathbb{R}^d)$ sense! A similar formula, *mutatis mutandis*, works for $\mathcal{F}^{-1}$.

## 1.9. Differential Geometry

[3] Schuller, "The Fourier Operator - L18 - Frederic Schuller".

**Definition 1.9.1 (Chart & Coordinates)**: Let $X$ be a topological space. A $d$-dimensional *chart* is a tuple $(U, \varphi)$ where $U \subseteq X$ is an open set in $X$ and $\varphi = (x^1, ..., x^d) : U \to \varphi(U)$ a homeomorphism onto its image $\varphi(U) \subseteq \mathbb{R}^d$ which is an open set of $\mathbb{R}^d$. The functions $x^i$ are called *coordinates*.

**Definition 1.9.2 (Transition Maps & Smoothly Compatible Charts)**: Let $(U_1, \varphi_1)$, $(U_2, \varphi_2)$ be two overlapping charts of a topological space $X$. One can show that overlapping charts must have the same dimension $d$. Consider the nonempty open intersection of their open sets $I = U_1 \cap U_2$, and the images of this intersection $\varphi_1(I), \varphi_2(I) \subseteq \mathbb{R}^d$. Their *transition maps* are the mappings

$$\varphi_2 \circ \varphi_1^{-1} : \varphi_1(I) \to \varphi_2(I)$$
$$\varphi_1 \circ \varphi_2^{-1} : \varphi_2(I) \to \varphi_1(I)$$

Two charts are *smoothly compatible* if their transition maps are smooth.

**Definition 1.9.3 (Atlas & Smooth Manifold)**: A *smooth atlas* of a topological space $X$ is a (possible uncountably infinite or even worse) collection of pairwise smoothly compatible charts that completely cover the space $X$.

A *smooth manifold* $M$ is a second countable Hausdorff topological space with a *smooth atlas*.

A $d$-dimensional smooth manifold is one where every chart of the atlas maps to an open subset of $d$-dimensional Euclidean space.

I suggest watching Frederick Schuller's video lecture[4] for an introduction to manifolds.

When we say "manifold" we almost always mean a "smooth $d$-dimensional manifold". When we are working with manifolds in practice we almost always have an atlas with a *finite* amount of charts, and sometimes we need only *one* chart.

**Definition 1.9.4 (Diffeomorphism & Diffeomorphic)**: Given two smooth manifolds $M$, $N$ a map $f : M \to N$ is called a *diffeomorphism* if it is bijective and both $f$ and its inverse $f^{-1}$ are smooth.

Two manifolds are called diffeomorphic if there exists a diffeomorphism between them.

In other words, a diffeomorphism is an *isomorphism* between smooth manifolds.

[4]Schuller, "Topological Manifolds and Manifold Bundles- Lec 06 - Frederic Schuller".

**Definition 1.9.5 (Manifold Embedding)**: Given two manifolds $M$, $N$ a map $f : M \to N$ is called a *embedding* if it is a diffeomorphism onto its image.

The above definition of a smooth manifold contains the technical terms *second countable* and *Hausdorff* because we then have the following very nice result.

**Theorem 1.9.1 (Whitney Embedding Theorem)**: Any smooth manifold (required to be Hausdorff and second-countable) can be smoothly embedded in $\mathbb{R}^d$ for some $d \in \mathbb{N}$.

Consider a smooth manifold $M$ and a differentiable curve $\gamma : \mathbb{R} \to M$. We want a notion of the "tangent vector" $\dot{\gamma}$ but the problem is that the standard definition

$$\dot{\gamma}(t) = \left(\frac{\partial}{\partial t}\gamma\right)(t) = \lim_{h \to 0} \frac{\gamma(t+h) - \gamma(t)}{h}$$

does **not** work because there is no notion of subtraction on a manifold. Instead, we do the following. We define the tangent vector $\dot{\gamma}(t) : C^1(M) \to \mathbb{R}$ as a linear operator on differentiable functions $f \in C^1(M)$ by:

$$(\dot{\gamma}(t))(f) = \frac{\partial}{\partial t}(f \circ \gamma) = \lim_{h \to 0} \frac{f(\gamma(t+h)) - f(\gamma(t))}{h}$$

This means that $\dot{\gamma}(t)$ eats a function and returns the directional derivative of that function in the direction of $\dot{\gamma}(t)$ at the point $\gamma(t) \in M$. We call $\dot{\gamma}(t)$ an element of the tangent space $T_{\gamma(t)}M$ at $\gamma(t)$.

Important and *characterizing* properties of a tangent vector $X_p = \dot{\gamma}(t)$ at $p = \gamma(t)$ are that for all $f, g \in C^1(M)$ and $a, b \in \mathbb{R}$:

- Linearity: $X_p(af + bg) = aX_p(f) + bX_p(g)$
- Product Rule: $X_p(f \cdot g) = X_p(f) \cdot g(p) + f(p) \cdot X_p(g)$

**Definition 1.9.6 (Tangent Vector & Space & Bundle)**: Let $M$ be a smooth manifold and $p \in M$ a point on the manifold. A linear operator $X_p : C^1(M) \to \mathbb{R}$ that satisfies the "product rule"

$$X_p(f \cdot g) = X_p(f) \cdot g(p) + f(p) \cdot X_p(g)$$

is called a *tangent vector* at $p$.

The collection of all tangent vectors at $p$ forms a vector space called the *tangent space* $T_pM$ of $p$.

The collection of all tangent spaces is called the *tangent bundle* $TM$.

This is fastest "workable" definition of a tangent vector. It is however *not* how I personally think of tangent vectors geometrically. Given that we can always embed a smooth man-

ifold smooth in Euclidean space one can **always** think of a tangent vector as a actual vector in Euclidean space.

I suggest watching Frederick Schuller's video lecture[5] for an introduction to tangent vector & spaces.

# 2. Geometric Image Processing

## 2.1. Gaussian Scale Space

**Definition 2.1.1 (Gaussian Kernel)**: The *Gaussian kernel* $G_s : \mathbb{R}^d \to \mathbb{R}_{>0}$ of *scale* $s > 0$ is defined as

$$G_s(x) = \frac{1}{(2\pi(2s))^{\frac{d}{2}}} e^{-\frac{1}{2}\frac{\|x\|^2}{2s}}. \tag{6}$$

When interpreted as a normal distribution, the variance $\sigma^2$ of the kernel can be read off as $\sigma^2 = 2s$. This also indicates, to me at least, that the wrong convention is used. We should have used a different scale parameter $s' = 2s$. Oh well...

**Definition 2.1.2 (Gaussian Scale Space)**: The *Gaussian scale space* $u(x, s) : \mathbb{R}^d \times \mathbb{R}_{>0} \to \mathbb{R}$ of a square integrable image $f : \mathbb{R}^d \to \mathbb{R}$ is defined as:

$$u(x, s) = (G_s * f)(x) = \int_{\mathbb{R}^d} G_s(x - y) f(y)\, \mathrm{d}y. \tag{7}$$

**Lemma 2.1.1 (Frequency Gaussian Kernel)**: The Gaussian kernel in the frequency domain is another Gaussian-like kernel:

$$\hat{G}_s(\omega) = (\mathcal{F}G_s)(\omega) = \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-s\|\omega\|^2} \tag{8}$$

**Lemma 2.1.2 (Semigroup Property of the Gaussian Kernel)**: For all $s, t > 0$ we have

$$G_s * G_t = G_{s+t}.$$

*Proof*: In the frequency domain is it readily checked that

$$(2\pi)^{\frac{d}{2}} \widehat{G_s} \cdot \widehat{G_t} = \hat{G}_{s+t} \tag{9}$$

Taking the inverse Fourier transform of this equation gives the result. □

**Definition 2.1.3 (Heat Equation)**: The *heat equation* is defined as the following PDE on $\mathbb{R}^d$

$$\begin{cases} \frac{\partial u}{\partial s}(x, s) = (\Delta u)(x, s) & \text{for } x \in \mathbb{R}^d, s > 0 \\ u(x, 0) = f(x) & \text{for } x \in \mathbb{R}^d \end{cases} \tag{10}$$

**Proposition 2.1.1**: The solution of the heat equation (10) is the Gaussian scale space (7).

*Proof*: By taking the Fourier transform of (10) (where we use (5)) we get the PDE

$$\begin{cases} \frac{\partial \hat{u}}{\partial s}(\omega, s) = -\|\omega\|^2 \hat{u}(\omega, s) & \text{for } \omega \in \mathbb{R}^d, s > 0 \\ \hat{u}(\omega, 0) = \hat{f}(\omega) & \text{for } \omega \in \mathbb{R}^d \end{cases}. \tag{11}$$

This PDE is quickly solved

$$\hat{u}(\omega, s) = e^{-s\|\omega\|^2} \hat{f}(\omega).$$

We see the Fourier transform of the Gaussian kernel (8) appearing

$$\ldots = (2\pi)^{\frac{d}{2}} (\mathcal{F}G_s)(\omega)(\mathcal{F}f)(\omega).$$

Applying the convolution theorem:

$$\ldots = (\mathcal{F}(G_s * f))(\omega).$$

Taking the inverse Fourier transform on both sides:

$$u(x, s) = (G_s * f)(x).$$

□

## 2.2. Poisson Scale Space

**Definition 2.2.1 (Poisson Kernel)**:

$$H_s(x) = \frac{1}{|S^d|} \frac{2s}{\left(s^2 + \|x\|^2\right)^{\frac{d+1}{2}}} \tag{12}$$

where $|S^d|$ denotes the surface area of the $d$ dimensional unit sphere embedded in $\mathbb{R}^{d+1}$.

$$S^d = \{x \in \mathbb{R}^{d+1} \mid \|x\| = 1\}$$

An explicit formula for this area is

$$|S^{d-1}| = \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)}.$$

[5] Schuller, "Differential Structures: The Pivotal Concept of Tangent Vector Spaces - Lec 09 - Frederic Schuller".

**Definition 2.2.2 (Poisson Scale Space)**: The *Poisson scale space* $u(x,s) : \mathbb{R}^d \times \mathbb{R}_{>0} \to \mathbb{R}$ of an image $f : \mathbb{R}^d \to \mathbb{R}$ is defined as:

$$u(x,s) = (H_s * f)(x) \tag{13}$$

where $H_s$ is the Poisson kernel (12).

**Definition 2.2.3 (Poisson PDE)**:

$$\begin{cases} \frac{\partial^2 u}{\partial s^2}(x,s) = (-\Delta u)(x,s) & \text{for } x \in \mathbb{R}^d, s > 0 \\ u(x,0) \quad = f(x) & \text{for } x \in \mathbb{R}^d \end{cases}. \tag{14}$$

The minus Laplacian $-\Delta$ (as seen in the Poisson PDE) in the frequency domain takes the form of a multiplication with $\|\omega\|^2$. This observation motivates us to define an operator, $\sqrt{-\Delta}$, that takes the form of a multiplication with $\|\omega\|$ in the frequency domain.

$$-\Delta f \longleftrightarrow \|\omega\|^2 \hat{f}, \quad \sqrt{-\Delta} f \longleftrightarrow \|\omega\| \hat{f}$$

Indeed, the square of this new operator would nicely correspond with $-\Delta$:

$$\sqrt{-\Delta}^2 = -\Delta$$

**Definition 2.2.4 (Square Root of Minus Laplacian)**: The operator $\sqrt{-\Delta}$ is defined through its action in the frequency domain:

$$\mathcal{F}\left(\sqrt{-\Delta} f\right)(\omega) = \|\omega\| (\mathcal{F} f)(\omega)$$

**Lemma 2.2.1 (Separation of Poisson PDE)**: Consider the Poisson PDE. We can write the PDE also as

$$(\partial_s^2 + \Delta) u = 0.$$

Using the square root of the minus Laplacian we can separate this as

$$(\partial_s^2 + \Delta) = \left(\partial_s + \sqrt{-\Delta}\right)\left(\partial_s - \sqrt{-\Delta}\right)$$

**Definition 2.2.5 (Half-Poisson PDE)**:

$$\begin{cases} \frac{\partial u}{\partial s}(x,s) = \left(-\sqrt{-\Delta} u\right)(x,s) & \text{for } x \in \mathbb{R}^d, s > 0 \\ u(x,0) \quad = f(x) & \text{for } x \in \mathbb{R}^d \end{cases} \tag{15}$$

**Proposition 2.2.1 (Stable Solutions of the Poisson PDEs)**: Show that the half-Poisson PDE (15) produces the solutions of the Poisson PDE (14) that do not blow up in the sense in the $\mathbb{L}^2(\mathbb{R}^d)$ sense.

*Proof*: Taking the Fourier transform of the Poisson PDE (14) gives:

$$\begin{cases} \frac{\partial^2 \hat{u}}{\partial s^2}(\omega,s) = \|\omega\|^2 \hat{u}(\omega,s) & \text{for } \omega \in \mathbb{R}^d, s > 0 \\ \hat{u}(\omega,0) \quad = \hat{f}(\omega) & \text{for } \omega \in \mathbb{R}^d \end{cases}. \tag{16}$$

This PDE is quickly solved

$$\hat{u}(\omega,s) = c_-(\omega) e^{-s\|\omega\|} \hat{f}(\omega) + c_+(\omega) e^{s\|\omega\|} \hat{f}(\omega),$$

with $c_-(\omega) + c_+(\omega) = 1$. We see that the second term of the solution blows up, so $c_+(\omega) = 0$ and $c_-(\omega) = 1$:

$$\hat{u}(\omega,s) = e^{-s\|\omega\|} \hat{f}(\omega).$$

Taking the Fourier transform of the half-Poisson PDE (15) gives:

$$\begin{cases} \frac{\partial \hat{u}}{\partial s}(\omega,s) = -\|\omega\| \hat{u}(\omega,s) & \text{for } \omega \in \mathbb{R}^d, s > 0 \\ \hat{u}(\omega,0) = \hat{f}(\omega) & \text{for } \omega \in \mathbb{R}^d \end{cases}.$$

The solution of this PDE is quickly seen to be

$$\hat{u}(\omega,s) = e^{-s\|\omega\|} \hat{f}(\omega).$$

So, indeed, the half-Poisson PDE (15) produces the solutions of the Poisson PDE (14) that do not blow up in the sense in the $\mathbb{L}^2(\mathbb{R}^d)$ sense. □

**Lemma 2.2.2 (Frequency Poisson Kernel)**: The Poisson kernel in the frequency domain is

$$\left(\widehat{H}_s\right)(\omega) = (\mathcal{F} H_s)(\omega) = \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-s\|\omega\|} \tag{17}$$

*Proof*: We only show for $d = 1$ that $\mathcal{F}^{-1}\widehat{H}_s = H_s$.

$$\left(\mathcal{F}^{-1}\widehat{H}_s\right)(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\omega x} \widehat{H}_s(\omega)\, \mathrm{d}\omega$$

Filling in $\widehat{H}_s$

$$\ldots = \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\omega x - s|\omega|}\, \mathrm{d}\omega,$$

we split the integral over $\mathbb{R}$ into a negative and positive part so we can rewrite $|\omega| \longrightarrow \pm\omega$

$$\begin{aligned} \ldots = &\frac{1}{2\pi} \int_{-\infty}^{0} e^{(ix+s)\omega}\, \mathrm{d}\omega \\ &+ \frac{1}{2\pi} \int_{0}^{\infty} e^{(ix-s)\omega}\, \mathrm{d}\omega, \end{aligned}$$

these two terms are then straightforwardly integrated:

$$\ldots = \frac{1}{2\pi} \frac{1}{ix+s} - \frac{1}{2\pi} \frac{1}{ix-s}$$

collecting everything:

$$\ldots = \frac{1}{2\pi}\frac{2s}{x^2+s^2}$$

which is indeed $H_s$. □

**Lemma 2.2.3 (Semigroup Property of the Gaussian Kernel)**: For all $s, t > 0$ we have

$$H_s * H_t = H_{s+t}.$$

*Proof*: In the frequency domain is it readily checked that

$$(2\pi)^{\frac{d}{2}}\widehat{H}_s \cdot \widehat{H}_t = \widehat{H}_{s+t}$$

Taking the inverse Fourier transform of this equation gives the result. □

**Proposition 2.2.2**: The solution of the half-Poisson PDE (and thus also the solutions of the Poisson PDE that do not blow up) is the Poisson scale space (13).

*Proof*: By taking the Fourier transform of the half-Poisson PDE (15) we get the PDE

$$\begin{cases} \frac{\partial \hat{u}}{\partial s}(\omega, s) = -\|\omega\|\hat{u}(\omega, s) & \text{for } \omega \in \mathbb{R}^d, s > 0 \\ \hat{u}(\omega, 0) = \hat{f}(\omega) & \text{for } \omega \in \mathbb{R}^d \end{cases}.$$

This PDE is quickly solved

$$\hat{u}(\omega, s) = e^{-s\|\omega\|}\hat{f}(\omega).$$

We see the Fourier transform of the Poisson kernel (17) appearing

$$\ldots = (2\pi)^{\frac{d}{2}}(\mathcal{F}H_s)(\omega)(\mathcal{F}f)(\omega).$$

Applying the convolution theorem:

$$\ldots = (\mathcal{F}(H_s * f))(\omega).$$

Taking the inverse Fourier transform on both sides:

$$u(x, s) = (H_s * f)(x).$$

□

## 2.3. Gaussian Scale Space on Bounded Domain

**Definition 2.3.1 (Heat Equation on Bounded Domain)**:

$$\begin{cases} \frac{\partial u}{\partial s}(x, s) = (\Delta u)(x, s) & \text{for } x \in \Omega, s > 0 \\ (\nabla u)(x, s) \cdot n(x) = 0 & \text{for } x \in \partial\Omega \\ u(x, 0) = f(x) & \text{for } x \in \Omega \end{cases} \tag{18}$$

where $\Omega = [0, \pi] \times [0, \pi] \subset \mathbb{R}^2$, $\partial\Omega$ indicates the boundary of $\omega$, and $n(x)$ is a normal vector of the boundary at $x$.

**Proposition 2.3.1 (Preservation of mass of Heat Equation)**: Consider the heat equation on the bounded domain (18). The total "mass" in the domain $\Omega$

$$\int_\Omega u$$

is constant.

*Proof*: We simply apply the divergence theorem and the boundary condition:

$$\frac{\partial}{\partial s}\int_\Omega u = \int_\Omega \Delta u = \int_\Omega \nabla \cdot \nabla u = \int_{\partial\Omega} \nabla u \cdot n = 0$$

□

**Proposition 2.3.2 (Solution of Heat Equation on Bounded Domain)**: A general solution is a superposition

$$u(x, y, s) = \sum_{m,n} A_{mn} f_{mn}(x, y, s)$$

where $A_{mn} \in \mathbb{R}$, and

$$f_{mn}(x, y, s) = \cos(mx)\cos(ny)e^{-(m^2+n^2)s}$$

*Proof*: Applying the method of separation, we hypothesize a solution of the form

$$f(x, y, s) = X(x)Y(y)S(s).$$

Applying the PDE we get

$$XYS' = X''YS + XY''S$$

we can rewrite this as

$$\frac{X''}{X} + \frac{Y''}{Y} = \frac{S'}{S}$$

this equation should be true for all $x, y, s$. This implies that all terms are constant, let us name the constants $\mu, \eta, \lambda$. We get a system of 3 ODEs:

$$\begin{cases} X'' = \mu X \\ Y'' = \eta Y \\ S' = \lambda S \end{cases}$$

with $\mu + \eta = \lambda$. We should not forget the boundary condition. Expressing this in terms of $X$ and $Y$ we get

$$X'(0) = X'(\pi) = 0,$$
$$Y'(0) = Y'(\pi) = 0.$$

Solving for $X$ and $Y$ we get

$$X_m(x) = \cos(mx),$$
$$Y_n(y) = \cos(ny)$$

for any $m, n \in \mathbb{N}$. So $\mu_m = -m^2$, $\eta_n = -n^2$, and $\lambda_{mn} = -(m^2 + n^2)$.

$$S_{mn}(s) = e^{-(m^2+n^2)s}$$

Thus, a specific solution is

$$f_{mn}(x, y, s) = \cos(mx)\cos(ny)e^{-(m^2+n^2)s}$$

□

**Definition 2.3.2 (Fractional Laplacian on Bounded Domain)**: We have seen above that

$$-\Delta f_{mn} = (m^2 + n^2) f_{mn}$$

We introduce a new operator, $-(-\Delta)^a$, defined by its action on $f_{mn}$:

$$-(-\Delta)^a f_{mn} = -(m^2 + n^2)^a f_{mn}$$

## 2.4. Gaussian Derivatives

**Definition 2.4.1 (Multi-index Derivative)**: Let $f : C^{\infty(\mathbb{R}^d)}$, and $\alpha \in \mathbb{N}^d$ be a *multi-index*. We define multi-index derivative $D^\alpha$ as

$$D^\alpha = \partial_1^{\alpha_1}\partial_2^{\alpha_2}\cdots\partial_d^{\alpha_d}$$

*Example 2.4.1*:

$$D^{1,2,1} f = \frac{\partial}{\partial x_1}\frac{\partial^2}{\partial x_2^2}\frac{\partial}{\partial x_3} f = \frac{\partial^4 f}{\partial x_1 \partial x_2^2 \partial x_3}$$

**Definition 2.4.2 (Gaussian Derivative)**: We define a *Gaussian derivative* $D_s^\alpha$ of a function $f : \mathbb{R}^d \to \mathbb{R}$ by first convolving with a Gaussian kernel $G_s$ (6) of a certain scale $s$ and then taking the $\alpha \in \mathbb{N}^d$ derivative (where $\alpha$ is a multi-index). Equivalently, this is the same as convolving with a derivative of a Gaussian.

$$D_s^\alpha f := D^\alpha(G_s * f) = (D^\alpha G_s) * f \tag{19}$$

**Lemma 2.4.1 (Gaussian Derivative in Frequency Domain)**: The Gaussian derivative $D_s^\alpha$ in the frequency domain is

$$\mathcal{F}(D_s^\alpha f) = (2\pi)^{\frac{d}{2}}(i\omega)^\alpha \hat{G}_s \hat{f}$$

*Proof*:

$$\begin{aligned}\mathcal{F}\left(\partial_j^s f\right) &= \mathcal{F}\left((\partial_j G_s) * f\right) \\ &= (2\pi)^{\frac{d}{2}}\mathcal{F}\left(\partial_j G_s\right) \cdot \hat{f} \\ &= (2\pi)^{\frac{d}{2}} i\omega_j \hat{G}_s \cdot \hat{f}\end{aligned}$$

□

**Lemma 2.4.2**: Higher order Gaussian derivatives can be expressed in terms of lower order Gaussian derivatives in the sense that

$$D_s^n = \left(D_{\frac{s}{n}}^1\right)^n$$

where for convenience we consider the $d = 1$ case.

*Proof*: In the frequency domain we have that

$$\begin{aligned}\mathcal{F}(D_s^n f) &= \sqrt{2\pi}(i\omega)^n \hat{G}_s \hat{f} \\ \mathcal{F}\left(D_{\frac{s}{n}}^1 f\right) &= \sqrt{2\pi}(i\omega)\hat{G}_{\frac{s}{n}} \hat{f}\end{aligned}$$

We see that

$$\mathcal{F}\left(\left(D_{\frac{s}{n}}^1\right)^n f\right) = \sqrt{2\pi}^n (i\omega)^n \left(\hat{G}_{\frac{s}{n}}\right)^n \hat{f}$$

using (9)

$$\sqrt{2\pi}\widehat{G}_s \cdot \widehat{G}_t = \hat{G}_{s+t}$$

So

$$\mathcal{F}\left(\left(D_{\frac{s}{n}}^1\right)^n f\right) = \sqrt{2\pi}(i\omega)^n \hat{G}_s \hat{f} = \mathcal{F}(D_s^n f)$$

from which we can conclude that, indeed,

$$D_s^n = \left(D_{\frac{s}{n}}^1\right)^n$$

□

**Proposition 2.4.1 (Derivative is Unbounded)**: Consider Schwartz space $S(\mathbb{R})$ equipped with the $\mathbb{L}^2$ norm. The derivative operator is unbounded in this space.

*Proof*: Consider $\varphi_{n(x)} = e^{-nx^2}$ for $n \in \mathbb{N}$. Indeed, $\varphi_n$ is in Schwartz space for every $n$. The derivative is $\varphi'_{n(x)} = -2nxe^{-nx^2}$. The $\mathbb{L}^2$ norm of $\varphi_n$ is

$$\|\varphi_n\|^2 = \int_{\mathbb{R}} e^{-2nx^2} = \frac{1}{\sqrt{2n}}\int_{\mathbb{R}} e^{-y^2}$$

where $y = \sqrt{2n}x$. The $\mathbb{L}^2$ norm of $\varphi'_n$ is

$$\|\varphi'_n\|^2 = \int_{\mathbb{R}} -4n^2x^2e^{-2nx^2} = \sqrt{2n}\int_{\mathbb{R}} y^2 e^{-y^2}$$

We see that

$$\frac{\|\varphi'_n\|}{\|\varphi\|} = 2nC$$

for some constant $C$. This grows without bound when we let $n \to \infty$, and thus the derivative operator is unbounded. □

**Proposition 2.4.2 (Gaussian Derivative is Bounded)**: Consider the space of square integrable functions $\mathbb{L}^2(\mathbb{R}^d)$. The Gaussian derivative $D_s^\alpha$ (19) is a bounded operator on this space.

*Proof*: We only show this for the one dimensional case $d = 1$. We are interested in the norm of $D_s f$. Since the Fourier transform is unitary we can equivalently analyze the Fourier transform of $D_s f$. We have already seen that

$$\mathcal{F}(D_s f) = \sqrt{2\pi} i\omega \hat{G}_s \cdot \hat{f}$$

Taking the norm of this gives

$$\|\mathcal{F}(D_s f)\|_2 \leq \left\|\sqrt{2\pi} i\omega \hat{G}_s\right\|_\infty \left\|\hat{f}\right\|_2$$

Now we have that

$$\left\|\sqrt{2\pi} i\omega \hat{G}_s\right\|_\infty = \left\|i\omega e^{-s\omega^2}\right\|_\infty = \frac{1}{\sqrt{2se}}$$

All in all

$$\|D_s f\| = \|\mathcal{F}(D_s f)\| \leq \frac{1}{\sqrt{2se}} \left\|\hat{f}\right\| = \frac{1}{\sqrt{2se}} \|f\|$$

and thus $D_s$ is a bounded operator. □

## 2.5. Hermite Polynomials

The *Hermite polynomials* $H_n(x)$ appear naturally when taking derivatives of Gaussians. Consider $G(x) = e^{-x^2}$. Taking derivatives:

$$\begin{aligned}
G(x) &= (1)G(x), \\
G'(x) &= (-2x)G(x), \\
G''(x) &= (4x^2 - 2)G(x), \\
G'''(x) &= (-8x^3 + 12x)G(x), \\
&\vdots
\end{aligned}$$

The Hermite polynomials are defined as the polynomials seen above, but then with (possibly) an extra minus sign to make the highest order term positive.

$$\begin{aligned}
H_0(x) &= 1, \\
H_1(x) &= 2x, \\
H_2(x) &= 4x^2 - 2, \\
H_3(x) &= 8x^3 - 12x.
\end{aligned}$$

**Definition 2.5.1 (Hermite Polynomials)**: Let $G(x) = e^{-x^2}$. The $n$th *Hermite polynomial* $H_n$ is defined as the polynomial that satisfies

$$G^{(n)}(x) = (-1)^n H_n(x) G(x). \tag{20}$$

**Proposition 2.5.1 (Recurrence for Hermite Polynomials)**:

$$H_{n+1}(x) = -H_n'(x) + 2xH_n(x)$$

*Proof*: We differentiate (20)

$$\begin{aligned}
G^{(n+1)}(x) = &(-1)^n H_n'(x) G(x) \\
&+(-1)^n H_n(x) G'(x),
\end{aligned}$$

using $G'(x) = -2xG(x)$ we see that

$$\begin{aligned}
&G^{(n+1)}(x) = \\
&(-1)^{n+1}(-H_n'(x) + 2xH_n(x))G(x),
\end{aligned}$$

and we read off that

$$H_{n+1}(x) = -H_n'(x) + 2xH_n(x).$$

□

**Proposition 2.5.2 (Gaussian Derivatives in Hermite Polynomials)**: Consider the Gaussian kernel $G_s$ of scale $s$ in one dimension $d = 1$. One has:

$$G_s^{(n)}(x) = (-1)^n \frac{1}{\sqrt{4s}^n} H_n\left(\frac{x}{\sqrt{4s}}\right) G_s(x)$$

*Proof*: We start by identifying that $G_s$ is just a rescaling of $G$

$$G_s(x) = \frac{1}{\sqrt{4\pi s}} e^{-\frac{x^2}{4s}} = \frac{1}{\sqrt{4\pi s}} G\left(\frac{x}{\sqrt{4s}}\right)$$

where $G(x) = e^{-x^2}$, as before. Taking the $n$'th derivative:

$$G_s^{(n)}(x) = \frac{1}{\sqrt{4\pi s}} \frac{1}{\sqrt{4s}^n} G^{(n)}\left(\frac{x}{\sqrt{4s}}\right)$$

plugging in $G^{(n)}$ (20):

$$G_s^{(n)}(x) = \frac{1}{\sqrt{4\pi s}} \frac{1}{\sqrt{4s}^n} (-1)^n H_n\left(\frac{x}{\sqrt{4s}}\right) G\left(\frac{x}{\sqrt{4s}}\right)$$

we identify $G_s$ again, and find:

$$G_s^{(n)}(x) = (-1)^n \frac{1}{\sqrt{4s}^n} H_n\left(\frac{x}{\sqrt{4s}}\right) G_s(x)$$

□

**Definition 2.5.2 (Hermite Basis)**: The following functions form an orthonormal basis for $\mathbb{L}^2(\mathbb{R})$

$$\eta_n(x) = \frac{1}{\sqrt{\sqrt{\pi} 2^n n!}} H_n(x) G\left(\frac{x}{\sqrt{2}}\right)$$

## 2.6. Separable and Isotropic Filters

**Definition 2.6.1 (Isotropic)**: A function $f : \mathbb{R}^d \to \mathbb{R}$ is called *isotropic* if there exists a $g : \mathbb{R}_{\geq 0} \to \mathbb{R}$ such that $f(x) = g(\|x\|)$.

**Proposition 2.6.1**: A function $f : \mathbb{R}^2 \to \mathbb{R}$ is isotropic if and only if

$$(-y\partial_x + x\partial_y)f = 0 \tag{21}$$

*Proof*: Let us proof this through a sequence of equivalences. We first turn to polar coordinates

$$f'(r, \theta) = f(r\cos\theta, r\sin\theta).$$

Isotropy means that

$$f(r\cos\theta, r\sin\theta) = g(r)$$

for some $g : \mathbb{R} \to \mathbb{R}$, but this equivalent to saying that

$$\frac{\partial f'}{\partial \theta} = 0.$$

Expanding the left hand side, we see that this is equivalent to

$$\begin{aligned}\frac{\partial f'}{\partial \theta} &= -r\sin\theta\frac{\partial f}{\partial x} + r\cos\theta\frac{\partial f}{\partial y} \\ &= -y\frac{\partial f}{\partial x} + x\frac{\partial f}{\partial y} = 0.\end{aligned}$$

We thus see that (21) holds if and only if $f$ is isotropic. □

**Definition 2.6.2 (Separable)**: A function $f : \mathbb{R}^d \to \mathbb{R}$ is called *separable* if there exists $f_1, f_2, ..., f_d : \mathbb{R} \to \mathbb{R}$ such that

$$f(x) = f_1(x_1) \cdot f_2(x_2) \cdot \cdots \cdot f_d(x_d)$$

**Proposition 2.6.2 (Separable Convolution)**: Let $f, g : \mathbb{R}^d \to \mathbb{R}$ be two function, and let $f$ be separable. A practical property of a separable function is that their convolution can also be separated:

$$f * g = f_1 *_1 f_2 *_2 \cdots * f_d *_d g$$

where $*_j$ is a convolution in only the $j$'th slot/dimension:

$$(f_j *_j g)(x) =$$

$$\int_{\mathbb{R}} f_j(x_j - y)g(x_1, ..., x_{j-1}, y, x_{j+1}, ..., x_d)\,\mathrm{d}y$$

*Proof*: For convenience let us only consider the $d = 2$ case.

$$\begin{aligned}&(f * g)(x, y) \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} f(x - x', y - y')g(x', y')\,\mathrm{d}x'\,\mathrm{d}y' \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} f_1(x - x')f_2(y - y')g(x', y')\,\mathrm{d}x'\,\mathrm{d}y' \\ &= \int_{\mathbb{R}} f_1(x - x')\left(\int_{\mathbb{R}} f_2(y - y')g(x', y')\,\mathrm{d}y'\right)\mathrm{d}x' \\ &= \int_{\mathbb{R}} f_1(x - x')(f_2 *_2 g)(x', y)\,\mathrm{d}x' \\ &= (f_1 *_1 (f_2 *_2 g))(x, y)\end{aligned}$$

And thus

$$f * g = f_1 *_1 f_2 *_2 g$$

□

**Proposition 2.6.3 (The only Seperable and Isotropic Filter)**: A differentiable, isotropic, and separable function $G : \mathbb{R}^d \to \mathbb{R}$ must be of the form

$$G(x) = Ae^{B\|x\|^2}$$

*Proof*: Being isotropic means that

$$G(x) = R\left(\|x\|^2\right),$$

for some $R : \mathbb{R}_{\geq 0} \to \mathbb{R}$. Being separable means that

$$G(x) = g_1(x_1) \cdot \cdots \cdot g_d(x_d).$$

Through rescaling we may assume without loss of generality that $G(0) = 1$, so we have

$$1 = R(0) = g_1(0)g_2(0)\cdots g_d(0).$$

Again, through rescaling we may further assume that

$$g_1(0) = \cdots = g_d(0) = 1.$$

By inspecting $G$ on a axis, that is we set all $x_j$ to zero except for, for example, $x_1$, we see from the previous two equations together that

$$\begin{aligned}G(x_1, 0, ..., 0) &= R(x_1^2) \\ &= g_1(x_1)g_2(0)\cdots g_d(0) \\ &= \frac{g_1(x_1)}{g_1(0)}R(0) \\ &= g_1(x_1),\end{aligned}$$

where we have used that $R(0) = g_1(0) = 1$. Now this is interesting, we see that in general

$$g_j(x_j) = R(x_j^2).$$

So all $g_j$ are in fact "the same". Plugging this back into the separability equation of $G$ we find

$$G(x) = R(x_1^2)\cdots R(x_n^2).$$

Using the isotropy on the l.h.s we see that

$$R(x_1^2 + \cdots + x_d^2) = R(x_1^2)\cdots R(x_n^2).$$

We recognize this as *Cauchy's exponential functional equation* of which the only differentiable solutions are

$$R(r) = e^{Br}$$

and thus we see that

$$G(x) = e^{B\|x\|^2}$$

□

## 2.7. B-Splines

**Definition 2.7.1 (B-Spline Basis Function)**: The zeroth order *B-Spline basis function* is defined by

$$B_0(x) = \begin{cases} 1 \text{ if } -\frac{1}{2} < x < \frac{1}{2} \\ 0 \text{ otherwise} \end{cases}$$

The $n$th order basis function is defined recursively through a convolution:

$$B_n = B_{n-1} * B_0$$

Note that the above definition is nonstandard, usually the basis functions are defined through the *Cox-de-Boor* recursion formula. More information of the link between these two definitions can be found at[6]. Most notably, the above definition hides the fact that all the basis functions are piecewise polynomial.

**Proposition 2.7.1**:

$$\int_{\mathbb{R}} B_n(x)\,\mathrm{d}x = 1$$

*Proof*: We see immediately that

$$\int_{\mathbb{R}} B_0(x)\,\mathrm{d}x = 1.$$

We will use this as the base for our induction. So, suppose that $\int_{\mathbb{R}} B_{n-1}(x)\,\mathrm{d}x = 1$ for $n \geq 1$, we will show then that $\int_{\mathbb{R}} B_n(x)\,\mathrm{d} = 1$.

$$\begin{aligned}\int_{\mathbb{R}} B_n(x)\,\mathrm{d} &= \int_{\mathbb{R}} (B_{n-1} * B_0)(x)\,\mathrm{d}x \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} B_{n-1}(x-y)B_0(y)\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_{\mathbb{R}}\left(\int_{\mathbb{R}} B_{n-1}(x-y)\,\mathrm{d}x\right)B_0(y)\,\mathrm{d}y \\ &= \int_{\mathbb{R}} B_0(y)\,\mathrm{d}y = 1\end{aligned}$$

□

**Definition 2.7.2 (Shift Operator)**: the *Shift Operator* $\mathcal{S}_k$ is defined

$$(\mathcal{S}_k(f))(x) = f(x-k)$$

**Lemma 2.7.1 (Shifts Commute with Convolutions)**:

$$\mathcal{S}_k(f * g) = (\mathcal{S}_k(f)) * g. \tag{22}$$

*Proof*: We start by applying the definition of the shift operator

$$(\mathcal{S}_k(f * g))(x) = (f * g)(x-k)$$

we then apply the definition of the convolution

$$\ldots = \int_{\mathbb{R}} f(x-k-y)g(y)\,\mathrm{d}y$$

we reintroduce the shift operator

$$\ldots = \int_{\mathbb{R}} (S_k f)(x-y)g(y)\,\mathrm{d}y$$

and we reintroduce the convolution

$$\ldots = (S_k(f) * g)(x).$$

□

**Proposition 2.7.2**:

$$\sum_{k\in\mathbb{Z}} \mathcal{S}_k(B_n) = 1$$

*Proof*: We immediately see that

$$\sum_{k\in\mathbb{Z}} \mathcal{S}_k(B_0) = 1$$

We will use this as the base for our induction. So, suppose that $\sum_{k\in\mathbb{Z}} \mathcal{S}_k(B_{n-1}) = 1$ for $n \geq 1$, we will show then that $\sum_{k\in\mathbb{Z}} \mathcal{S}_k(B_n) = 1$.

[6] Johnson, "Relation between Cox-Deboor Recursion and Convolution (B-Spline Basis)".

$$\begin{aligned}\sum_{k\in\mathbb{Z}} \mathcal{S}_k(B_n) &= \sum_{k\in\mathbb{Z}} \mathcal{S}_k(B_{n-1} * B_0) \\ &= \sum_{k\in\mathbb{Z}} (\mathcal{S}_k(B_{n-1})) * B_0 \\ &= \left(\sum_{k\in\mathbb{Z}} \mathcal{S}_k(B_{n-1})\right) * B_0 \\ &= 1 * B_0 = 1\end{aligned}$$

Where we have used that shifts commute with convolutions (22). □

**Definition 2.7.3 (B-Spline Basis)**: A collection of scaled and shifted versions of $B_n$ is called a *B-Spline basis*.

## 2.8. Linear Least Squares

Suppose we have a *overdetermined* system of equations

$$Ax = y$$

where $A \in \mathbb{R}^{m\times n}, y \in \mathbb{R}^m$ are given, and we are searching for a $x \in \mathbb{R}^n$, with $m > n$. In general, this system does *not* necessarily have a solution because it is overdetermined. In this case we settle for the next best thing, the solution that is a close as possible.

**Definition 2.8.1 (Linear Least Squares)**: Let $A \in \mathbb{R}^{m\times n}, y \in \mathbb{R}^m$. We are looking for $x \in \mathbb{R}^n$ such that

$$x_{opt} = \underset{x\in\mathbb{R}^n}{\operatorname{argmin}} \|Ax - y\|^2. \tag{23}$$

This minimization problem is called *linear least squares*.

*Example 2.8.1 (Fitting a Line)*: Suppose we have 3 data points $(1,1),(2,3),(3,2)$ and we want to fit a line $l(x) = a + bx$ through these data points. In a sense we want to solve the overdetermined system

$$\begin{cases} a + b1 = 1 \\ a + b2 = 3 \\ a + b3 = 2 \end{cases}$$

or in matrix form

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}.$$

**Proposition 2.8.1 (Solution to Linear Least Squares)**: The solution to linear least squares (23) is

$$x_{opt} = (A^T A)^{-1} A^T y$$

where we have *assumed* that $A^T A$ is invertible, which is the case if the columns of $A$ are independent.

*Proof*: We define $f(x) = \|Ax - y\|^2$. Expanding this gives

$$\begin{aligned} f(x) &= (Ax - y)^T (Ax - y) \\ &= x^T A^T Ax - 2y^T Ax + y^T y \end{aligned}$$

The gradient of this is

$$(\nabla f)(x) = 2A^T Ax - 2A^T y$$

where we have used the facts that the gradient of $f(x) = y^T x$ is $(\nabla f)(x) = y$, and the gradient of $f(x) = x^T Sx$ is $(\nabla f)(x) = 2Sx$ for symmetric matrices $S$. We have as a necessary condition that

$$(\nabla f)(x_{opt}) = 0$$

Plugging in the gradient gives

$$A^T Ax_{opt} = A^T y$$

and thus

$$x_{opt} = (A^T A)^{-1} A^T y$$

□

**Proposition 2.8.2 (Least Squares with Regularization)**: Consider the an optimization problem where the objective is

$$\begin{aligned} x_{opt} &= \underset{x\in\mathbb{R}^n}{\operatorname{argmin}} f(x) \\ f(x) &= \|Bx - y\|^2 + \lambda x^T Bx + \mu x^T x. \end{aligned} \tag{24}$$

where $\lambda, \mu > 0$. The solution is

$$x_{opt} = (A^T A + \lambda B + \mu I)^{-1} B^T y.$$

*Proof*:

$$(\nabla f)(x) = 2A^T Ax - 2A^T y + 2\lambda Bx + 2\mu I$$

setting this equal to zero gives the result. □

**Proposition 2.8.3**: Let $A \in \mathbb{R}^{m\times n}$ be any matrix. The square matrix $M = A^T A + \mu I$ is invertible for any $\mu > 0$.

*Proof*: Suppose for the sake of contradiction that $M$ is not invertible, then their exists a $x \neq 0$ such that $Mx = 0$, more specifically we have that $x^T Mx = 0$. However, if we expand this product we get

$$\begin{aligned} 0 &= x^T Mx \\ &= x^T A^T Ax + x^T \mu I x \\ &= \|Ax\|^2 + \mu \|x\|^2 \\ &> 0 \end{aligned}$$

where the last inequality holds because $\mu > 0$. A contradiction appeared, and thus $M$ is invertible. □

*Proof*: Another proof goes by construction. $A^TA$ is a symmetric positive semidefinite matrix and is thus orthogonally diagonalizable $A^TA = Q\Lambda Q^T$, with $\Lambda$ a diagonal matrix of non-negative eigenvalues, and $Q^TQ = QQ^T = I$. We can rewrite $M$ as $M = Q(\Lambda + \mu I)Q^T$. The explicit inverse is then simply $M^{-1} = Q(\Lambda + \mu I)^{-1}Q^T$, where indeed $\Lambda + \mu I$ is invertible because all entries are positive because $\mu > 0$. □

## 2.9. Operator Exponential, One Parameter Semi-Groups, and Generators

**Definition 2.9.1 (Operator Exponential)**: Let $X$ be a Banach space. For a bounded linear operator $A : X \to X$ we define the operator exponential $e^A \in B(X)$

$$e^A = \sum_{n=0}^{\infty} \frac{1}{n!} A^n = I + A + \frac{1}{2}A^2 + \cdots$$

**Proposition 2.9.1 (Operator Exponential is Well-Defined)**: The operator exponential is a well-defined bounded linear operator.

*Proof*: We first clarify the definition. We really mean with $e^A$ the operator that maps $x \in X$ as follows

$$e^A x = \sum_{n=0}^{\infty} \frac{1}{n!} A^n x \tag{25}$$

The question is if this series indeed converges for all $x \in X$. To proof this we will use the fact that any absolutely convergent series in a Banach space is convergent. To this end, consider the series

$$\sum_{n=0}^{\infty} \left\| \frac{1}{n!} A^n x \right\| \tag{26}$$

if this converges in $\mathbb{R}$ the original series is, by definition, absolutely convergent, and thus the original series converges. Since $A$ was assumed to be a bounded linear operator, there exists a $M \geq 0$ such that $\|Ax\| \leq M\|x\|$. With this we can bound

$$\sum_{n=0}^{\infty} \left\| \frac{1}{n!} A^n x \right\| \leq \sum_{n=0}^{\infty} \frac{1}{n!} M^n \|x\| = e^M \|x\|$$

So, by the monotone convergence theorem, (26) converges, and thus (25) converges. From this we also see that the operator exponential is a bounded linear operator as we can bound it with $e^M$. □

The operator exponential pops up whenever one solves a linear first order ODE

**Proposition 2.9.2**: Let $X$ be a Banach space, and $A : X \to X$ a bounded linear operator. Consider the linear first order ODE

$$y'(t) = Ay(t)$$

where $y(t) \in B$. The solution of this ODE is

$$y(t) = e^{At} y(0),$$

We call $e^{At}$ the *solution operator* of the ODE.

*Proof*:

$$\begin{aligned}
\frac{\partial}{\partial t} e^{At} &= \frac{\partial}{\partial t}\left( I + At + \frac{1}{2}A^2t^2 + \frac{1}{6}A^3t^3 + \cdots \right) \\
&= A + A^2 t + \frac{1}{2}A^3 t^2 + \cdots \\
&= A\left( I + At + \frac{1}{2}A^2t^2 + \cdots \right) \\
&= Ae^{At}
\end{aligned}$$

which implies

$$y'(t) = \frac{\partial}{\partial t} e^{At} y(0) = Ae^{At} y(0) = Ay(t)$$

so indeed the solution is $y(t) = e^{At}y(0)$. □

*Example 2.9.1*: The solution of

$$y'(t) = ay(t)$$

with $a, y(t) \in \mathbb{R}$, is $y(t) = e^{at}y(0)$.

Another example is

$$y'(t) = Ay(t)$$

with $a \in \mathbb{R}^{d\times d}$ and $y \in \mathbb{R}^d$. The solution of this ODE is $y(t) = e^{At}y(0)$, where the exponential is the matrix exponential.

**Definition 2.9.2 (One-Parameter (Semi-)Group)**: A *one-parameter semi-group* is a set of elements $T_t$ indexed by $t \in \mathbb{R}_{\geq 0}$ with some operation $\star$ that satisfies the *semi-group property*

$$T_s \star T_t = T_{s+t} \text{ for all } s, t \geq 0.$$

The semigroup property implies that $\star$ is associative, commutative, and has the identity element $T_0$. If one can extend the definition from just $\mathbb{R}_{\geq 0}$ to the whole of $\mathbb{R}$, it is called a *one-parameter group*.

*Example 2.9.2*: We have already seen an example of a one-parameter semigroup, namely the solution operator $T_t = e^{At}$. Intuitivily, this should form a semigroup because running an ODE for $s$ time and then $t$ time is the same as running it for

$s+t$ time. We can also check algebraically from the definition that, indeed, $e^{As}e^{At} = e^{A(s+t)}$

**Lemma 2.9.1**: The operator exponential forms a one-parameter group in the sense that

$$e^{As}e^{At} = e^{A(s+t)}$$

*Proof*: Given that all infinite sums that we encounter below converge absolutely we may freely manipulate them without worries.

$$e^{A(s+t)} = \sum_{n=0}^{\infty} \frac{1}{n!}(A(s+t))^n$$

Binomial formula

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{i=0}^{n} \binom{n}{i} (As)^i (At)^{n-i}$$

Expand binomial coefficient

$$= \sum_{n=0}^{\infty} \sum_{i=0}^{n} \frac{1}{i!}(As)^i \frac{1}{(n-i)!}(At)^{n-i}$$

rewrite sum

$$= \sum_{m,n=0}^{\infty} \frac{1}{m!}(As)^m \frac{1}{n!}(At)^n$$

Sum over all possible multiplications is the same as the multiplication of two sums

$$= \left(\sum_{n=0}^{\infty} \frac{1}{n!}(As)^n\right)\left(\sum_{n=0}^{\infty} \frac{1}{n!}(At)^n\right)$$

$$= e^{As}e^{At}$$

□

*Corollary 2.9.1*: By plugging in $t = -s$ in the above lemma we get $e^{As}e^{-As} = e^0 = I$. This means that $\left(e^{As}\right)^{-1} = e^{-As}$

**Definition 2.9.3 (Generator)**: Let $X$ be a Banach space, and $T_t : X \to X$ one-parameter (semi-)group of bounded linear operators. If there exists a $A : X \to X$ such that

$$T_t = e^{At}$$

then $A$ is called the *generator* of the one-parameter (semi-)group.

**Lemma 2.9.2**: Let $X$ be a Banach space, and $T_t : X \to X$ a one-parameter (semi-)group of bounded linear operators with generator $A : X \to X$. One can find the generator through

$$A = \left.\frac{\partial}{\partial t}\right|_{t=0} T_t$$

*Proof*:

$$\left.\frac{\partial}{\partial t}\right|_{t=0} T_t = \left.\frac{\partial}{\partial t}\right|_{t=0} e^{At} = \left[Ae^{At}\right]_{t=0} = A$$

□

The above lemma suggests the following strategy to see if a one-parameter family $T_t$ is indeed generated by a single operator $A$:

1. Calculate $A = \left.\frac{\partial}{\partial t}\right|_{t=0} T_t$
2. Check if $e^{At} = T_t$

**Proposition 2.9.3 (Rotation Generator)**: Consider the one-parameter family of 2D rotation matrices $R_\theta$. One has

$$\left.\frac{\partial}{\partial \theta}\right|_{\theta=0} R_\theta = X \quad \text{and } e^{X\theta} = R_\theta.$$

where

$$X = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

*Proof*:

$$\frac{\partial}{\partial \theta} R_\theta = \begin{pmatrix} -\sin\theta & -\cos\theta \\ \cos\theta & -\sin\theta \end{pmatrix}$$

plugging in $\theta = 0$ gives the first result. For the second result, we simply calculate the matrix exponential and use our knowledge of the Taylor series of cos and sin:

$$\begin{aligned} e^{X\theta} &= I + X\theta + \frac{1}{2}(X\theta)^2 + \frac{1}{6}(X\theta)^3 \cdots \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\theta + \frac{1}{2}\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}\theta^2 + \cdots \\ &= \begin{pmatrix} 1 - \frac{\theta^2}{2} + \cdots & -\theta + \frac{\theta^3}{6} + \cdots \\ \theta - \frac{\theta^3}{6} + \cdots & 1 - \frac{\theta^2}{2} + \cdots \end{pmatrix} \\ &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \end{aligned}$$

□

The above proposition show that $X$ is the generator of rotations $R_\theta$ in $\mathbb{R}^2$

## 2.10. Roto-Translations

**Definition 2.10.1 (Roto-Translation Group)**: Let $v \in \mathbb{R}^d$ be a translation vector, and $R \in \mathrm{SO}(d) \subset \mathbb{R}^{d\times d}$ a rotation matrix. We let a *roto-translation* $(v, R)$ act $\triangleright$ on $\mathbb{R}^d$ in the obvious way:

$$(v, R) \triangleright x = Rx + v \tag{27}$$

We first rotate with $R$ and then translates with $v$.

We define the $d$-dimensional *roto-translation group* $\mathrm{SE}(d) = \mathbb{R}^d \rtimes \mathrm{SO}(d)$, also known as the *special euclidean* group, as the collection of all roto-translations with the group operation $\cdot$ being

$$(v_2, R_2) \cdot (v_1, R_1) = (R_2 v_1 + v_2, R_2 R_1) \tag{28}$$

The ordering of the subscripts in (28) is such because we imagine first performing $(v_1, R_1)$ and afterwards performing $(v_2, R_2)$. The group operation is natural as it plays nicely with the action $\triangleright$ (27):

$$g_2 \triangleright (g_1 \triangleright x) = (g_2 \cdot g_1) \triangleright x$$

**Definition 2.10.2 (Coordinates on Two Dimensional Roto-Translation Group)**: Consider the 2-dimensional roto-translation group $\mathrm{SE}(2)$. We use coordinates $(x, y, \theta) \in \mathbb{R}^2 \times [0, 2\pi)$ and group product

$$(x, y, \theta)(x', y', \theta') = \begin{pmatrix} x + x'\cos\theta - y'\sin\theta \\ y + x'\sin\theta + y'\cos\theta \\ \theta + \theta' \bmod 2\pi \end{pmatrix}$$

In these coordinates the identity element is $e = (0, 0, 0)$.

**Definition 2.10.3 (Roto-Translation Operator)**: Let $(v, R)$ be a roto-translation in $\mathbb{R}^d$. We define the *roto-translation operator* $\mathcal{U}_{v,R}$ by its action on scalar functions $f : \mathbb{R}^d \to \mathbb{C}$:

$$(\mathcal{U}_{v,R} f)(x) = f(R^{-1}(x - v))$$

or equivalently using the action $\triangleright$ of $\mathrm{SE}(d)$ on $\mathbb{R}^d$:

$$\mathcal{U}_{v,R} f = f \circ (g^{-1} \triangleright)$$

Intuitively, this simply first rotates the function with $R$ and then translates it with $v$. (Yes, really!)

**Proposition 2.10.1 (Translation Generator)**: Let $v \in \mathbb{R}^2$ and $f \in C^1(\mathbb{R}^2)$.

$$\frac{\partial}{\partial h}\bigg|_{h=0} (\mathcal{U}_{hv,I} f) = -(v_x \partial_x + v_y \partial_y) f$$

*Proof*: First, by definition we have

$$(\mathcal{U}_{hv,I} f)(x, y) = f(x - hv_x, y - hv_y)$$

Taking the $\frac{\partial}{\partial h}$ derivative on both sides, and expanding the r.h.s gives

$$\begin{aligned} \frac{\partial}{\partial h}(\text{r.h.s}) &= -v_x \frac{\partial f}{\partial x}(\ldots) - v_y \frac{\partial f}{\partial x}(\ldots) \\ &= \left(-(v_x \partial_x + v_y \partial_y) f\right)(\ldots) \end{aligned}$$

Plugging in $h = 0$ gives the result. □

**Proposition 2.10.2**: Consider an analytic function $f : \mathbb{R}^2 \to \mathbb{R}^2$. One has

$$e^{-t\partial_x} f = \mathcal{U}_{(t,0),I} f$$

*Proof*: We first expand the l.h.s to be

$$\begin{aligned} (e^{-t\partial_x} f)(x) &= \left(\left(\sum_{n=0}^{\infty} \frac{1}{n!}(-t\partial_x)^n\right) f\right)(x) \\ &= \sum_{n=0}^{\infty} \frac{f^{(n)}(x)}{n!}(-t)^n \end{aligned} \tag{29}$$

We recognize this last term as the Taylor expansion of $f$ around $x$ evaluated at $x - t$:

$$f(x - t) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x)}{n!}(-t)^n$$

The Taylor series converges everywhere as $f$ is assumed to be analytic. The result follows. □

The above two propositions show that $v_x \partial_x + v_y \partial_y$ is the generator of translations $\mathcal{U}_{v,I}$ of functions in $\mathbb{R}^2$.

**Proposition 2.10.3 (Rotation Generator)**: Let $R_\alpha \in \mathbb{R}^{2\times 2}$ be the two dimensional rotation matrix which rotates anti-clockwise with $\alpha \in \mathbb{R}$, and $f \in C^1(\mathbb{R}^2)$

$$\frac{\partial}{\partial \alpha}\bigg|_{\alpha=0} \left(\mathcal{U}_{0,R_\alpha} f\right) = -(-y\partial_x + x\partial_y) f$$

*Proof*: It is best to first translate the problem to polar coordinates $(r, \theta)$.

$$\left(\mathcal{U}_{0,R_\alpha} f\right)(r, \theta) = f(r, \theta - \alpha)$$

Taking the $\frac{\partial}{\partial \alpha}$ derivative and evaluating at $\alpha = 0$ gives

$$\frac{\partial}{\partial \alpha}\bigg|_{\alpha=0} \left(\mathcal{U}_{0,R_\alpha} f\right) = -\partial_\theta f$$

Knowing that $\partial_\theta = -y\partial_x + x\partial_y$ gives the result. □

*Proof*: By definition

$$\big(\mathcal{U}_{0,R_\alpha} f\big)(x,y) =$$
$$f(x\cos\alpha + y\sin\alpha, -x\sin\alpha + y\cos\alpha)$$

taking the $\frac{\partial}{\partial\alpha}$ derivative on both sides, and expanding the r.h.s gives

$$\frac{\partial}{\partial\alpha}(\text{r.h.s}) = (-x\sin\alpha + y\cos\alpha)\frac{\partial f}{\partial x}(\ldots) + (-x\cos\alpha - y\sin\alpha)\frac{\partial f}{\partial y}(\ldots)$$

plugging in $\alpha = 0$

$$= y\frac{\partial f}{\partial x}(x,y) - x\frac{\partial f}{\partial y}(x,y)$$
$$= \big(-(-y\partial_x + x\partial_y)f\big)(x,y)$$

□

**Proposition 2.10.4**: Consider an analytic function $f : \mathbb{R}^2 \to \mathbb{R}^2$. One has

$$e^{-\alpha(-y\partial_x + x\partial_y)} f = \mathcal{U}_{0,R_\alpha} f$$

*Proof*: It is best to first translate the equation to polar coordinates.

$$\big(e^{-\alpha\partial_\theta} f\big)(r,\theta) = f(r,\theta-\alpha)$$

The proof is then exactly similar to (29). □

The above two propositions show that $-y\partial_x + x\partial_y$ is the generator of rotations $\mathcal{U}_{0,R_\theta}$ of functions in $\mathbb{R}^2$.

## 2.11. Lie Group Theory

**Definition 2.11.1 (Lie Group)**: A Lie group is a group that is also a finite-dimensional smooth manifold, in which the group operations of multiplication and inversion are smooth maps.

I suggest watching Frederick Schuller's video lecture[7] for an introduction to Lie groups.

The tangent space $T_eG$ at the identity of a Lie group $G$ is special: it represents generators of one-parameter groups.

**Definition 2.11.2 (Lie Algebra of a Lie group)**: Let $G$ be a Lie group. The tangent space $T_eG$ at the identity is called the *Lie algebra* of the Lie group, and is denoted with $\mathfrak{g}$.

To go from the generators $\mathfrak{g}$ to the one-parameter subgroups we use the Lie group exponential.

**Definition 2.11.3 (Lie Group Exponential Curves and Map)**: Let $G$ be a lie group, and $X \in \mathfrak{g}$ a tangent vector. The *exponential curve* $\gamma(t) : \mathbb{R} \to G$ corresponding to $X$ is defined as the unique curve satisfying

$$\dot\gamma(0) = X \quad \text{and } \gamma(s)\gamma(t) = \gamma(s+t). \tag{30}$$

That means the exponential curve is a one-parameter group, with its generator being $X$. The *exponential* of $X$ is defined as

$$\exp(X) = \gamma(1)$$

this creates the *exponential map* $\exp : \mathfrak{g} \to G$.

**Lemma 2.11.1**: Let $G$ be a Lie group, $X \in \mathfrak{g}$ a tangent vector, and $\gamma(t) : \mathbb{R} \to G$ the corresponding exponential curve. Consider the curve

$$\eta(t) = \gamma(\lambda t) \tag{31}$$

This is the exponential curve of $\lambda X \in \mathfrak{g}$.

*Proof*: We show that $\eta$ satisfies the definition (30).

$$\dot\eta(t) = \lambda\dot\gamma(\lambda t) \Rightarrow \dot\eta(0) = \lambda\dot\gamma(0) = \lambda X$$

and

$$\eta(s)\eta(t) = \gamma(\lambda s)\gamma(\lambda t) = \gamma(\lambda(s+t)) = \eta(s+t)$$

□

*Corollary 2.11.1*: By plugging in $t = 1$ into (31) we get

$$\exp(\lambda X) = \eta(1) = \gamma(\lambda)$$

Which means the exponential curve $\gamma$ of $X$ can be written as

$$\gamma(t) = \exp(tX)$$

**Proposition 2.11.1 (Lie Group exponential reduces to Known Exponentials)**: Let $G$ be a Lie group where both the group and the Lie algebra $\mathfrak{g}$ consist of bounded linear operators. Then the Lie group exponential corresponds to the operator exponential.

If the Lie group and algebra consist of matrices then the Lie group exponential is just the matrix exponential.

**Definition 2.11.4 (Push-Forward)**: Let $M$, $N$ be two differentiable manifolds, and let $X \in T_pM$ be a tangent vector at $p \in M$. Let $\varphi : M \to N$ be a differentiable map. We define the *push-forward* $\varphi_* X \in T_{\varphi(p)}G$ of $X$ by $\varphi$ by its action on functions $f \in C^1(N)$:

$$(\varphi_* X f) = X(f \circ \varphi)$$

[7] Schuller, "Lie Groups and Their Lie Algebras - Lec 13 - Frederic Schuller".

**Definition 2.11.5 (Vector Field)**: Let $M$ be a differentiable manifold. A *vector field* $\mathcal{X} : M \to TM$ is an assignment of a tangent vector $X \in T_pM$ at every point $p \in M$ on the manifold.

The set of all smooth vector fields is denoted by $\Gamma(TM)$.

We often interpret a vector field as a linear operator $\mathcal{X} : C^\infty(M) \to C^\infty(M)$, just as we interpret tangent vector withs operator $X : C^1(M) \to \mathbb{R}$.

**Definition 2.11.6 (Left-Invariant Vector Field)**: Let $G$ be a Lie group. A vector field $\mathcal{X} \in \Gamma(TG)$ is called *left-invariant* if

$$(L_g)_* \mathcal{X}_h = \mathcal{X}_{gh} \text{ for all } g, h \in G$$

**Lemma 2.11.2 (Lie Algebra and Left-Invariant Vector Fields)**: Let $G$ be a Lie group. Every $X \in \mathfrak{g} = T_eG$ corresponds to a left-invariant vector field $\mathcal{X} \in \Gamma(TM)$, and vice versa

*Proof*: $\mathcal{X}_g = (L_g)_* X \longleftrightarrow X = \mathcal{X}_e$ □

**Proposition 2.11.2 (Left-Invariant Vector Field Generates Right-Translations)**: Let $G$ be a Lie group, $X \in \mathfrak{g}$ a tangent vector at the identity, $\mathcal{X} \in \Gamma(TM)$ the corresponding left-invariant vector field, and $\gamma : \mathbb{R} \to G$ the exponential curve of $X$. Consider the one parameter group $\mathcal{R}_{\gamma(t)}$. One has

$$\left.\frac{\partial}{\partial t}\right|_{t=0} \mathcal{R}_{\gamma(t)} = \mathcal{X}$$

*Proof*: Let $f \in C^1(G)$ be some dummy function, and $h \in G$ some dummy group element.

$$\begin{aligned}\frac{\partial}{\partial t}\Big(\mathcal{R}_{\gamma(t)} f\Big)(h) &= \frac{\partial}{\partial t} f(h\gamma(t)) = \frac{\partial}{\partial t}(f \circ L_h)(\gamma(t)) \\ &= \dot\gamma(t)(f \circ L_h) = \Big((L_h)_* \dot\gamma(t)\Big) f\end{aligned}$$

plugging in $t = 0$ gives

$$\left.\frac{\partial}{\partial t}\right|_{t=0} \Big(R_{\gamma(t)} f\Big)(h) = \Big((L_h)_* X\Big) f = \mathcal{X}_h f$$

□

*Corollary 2.11.2*: Taking the operator exponential, writing $\gamma(t) = e^{tX}$, and $\mathcal{X} = \mathcal{R}_X$, we can write the lemma also as

$$e^{t\mathcal{R}_X} = \mathcal{R}_{e^{tX}}$$

Notice that this formula contains two different exponentials: the Lie group exponential and the operator exponential.

**Proposition 2.11.3**: Let $G$ be a Lie group AND $\mathcal{X} \in \Gamma(TM)$ a left-invariant vector field. We have that $\mathcal{X}$ commutes with left-translations $\mathcal{L}_g$, i.e. is an *equivariant* operator.

$$\mathcal{X} \circ \mathcal{L}_g = \mathcal{L}_g \circ \mathcal{X}$$

*Proof*: Let $f \in C^1(G)$ be some dummy function, and $h \in G$ some dummy group element.

$$\begin{aligned}X_h(\mathcal{L}_g f) = X_h(f \circ L_{g^{-1}}) &= \Big((L_{g^{-1}})_* X_h\Big) f \\ &= X_{g^{-1}h} f = (Xf)(g^{-1}h) \\ &= (\mathcal{L}_g(Xf))(h),\end{aligned}$$

where the second equality is by definition of the pushforward, and the third because $\mathcal{X}$ is left-invariant. □

**Proposition 2.11.4**: Consider the Lie group $G = \mathrm{SE}(2)$ with coordinates $(x, y, \theta)$. Let $A_1 = \partial_x|_e$, $A_2 = \partial_y|_e$, $A_3 = \partial_\theta|_e$. Consider a $(x, y, \theta) = g \in G$ then

$$\mathcal{A}_1|_g = (L_g)_* A_1 = +\cos\theta \partial_x|_g + \sin\theta \partial_y|_g$$

$$\mathcal{A}_2|_g = (L_g)_* A_2 = -\sin\theta \partial_x|_g + \cos\theta \partial_y|_g$$

$$\mathcal{A}_3|_g = (L_g)_* A_3 = \partial_\theta|_g$$

*Proof*: Consider a dummy function $f \in C^1(G)$.

$$\begin{aligned}\Big((L_g)_* A_1\Big)(f) &= \partial_{x'}|_e (f \circ L_g)(x', y', \theta') \\ &= \partial_{x'}|_e f((x, y, \theta)(x', y', \theta'))\end{aligned}$$

Plugging in the group product and simply chugging along (using the chain rule) gives:

$$\ldots = +\cos\theta \partial_x|_g f + \sin\theta \partial_y|_g f$$

We derive that

$$\mathcal{A}_1 = +\cos\theta \partial_x + \sin\theta \partial_y$$

We notice that we never really needed the dummy function to make sense of the derivation; we simply need to know how the group product transform derivatives on the right. So, we simply take the group product as above and take the derivatives w.r.t $x'$, $y'$, and $\theta'$ at $x' = y' = \theta' = 0$:

$$\frac{\partial}{\partial x'}(\ldots) = (\cos\theta, \sin\theta, 0)$$

$$\frac{\partial}{\partial y'}(\ldots) = (-\sin\theta, \cos\theta, 0)$$

$$\frac{\partial}{\partial \theta'}(\ldots) = (0, 0, 1)$$

where $\ldots$ is the group product, and then we simply read of that

$$\mathcal{A}_1 = +\cos\theta\partial_x + \sin\theta\partial_y$$
$$\mathcal{A}_2 = \sin\theta\partial_x + \cos\theta\partial_y$$
$$\mathcal{A}_3 = \partial_\theta$$

□

## 2.12. Linear Constrained Optimization & Lagrange Multipliers

**Definition 2.12.1 (Linear Constrained Optimization)**: Let $S \in \mathbb{R}^{n\times n}$ be a symmetric positive semi-definite matrix. We define the following constrained optimization problem

$$x_* = \underset{\|x\|=1}{\operatorname{argmin}}\, x^T S x$$

**Proposition 2.12.1**: $x_*$ is the (normalized) eigenvector belonging to the smallest (in magnitude) eigenvalue $\lambda$ of $S$.

*Proof*: We prove using Lagrange multipliers. Define

$$\mathcal{L} = x^T S x + \lambda\left(1 - \|x\|^2\right)$$

Taking gradients

$$\frac{\partial \mathcal{L}}{\partial x} = 2Sx - 2\lambda x$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = 1 - \|x\|^2$$

Setting both to zero and solving gives the equations

$$Sx = \lambda x \quad \text{and} \quad \|x\| = 1$$

We see that $x$ is a (normalized) eigenvector. Plugging this back into the minimization problem

$$x_* = \underset{Sx=\lambda x, \|x\|=1}{\operatorname{argmin}}\, \lambda$$

and we see that $\lambda$ must be the smallest eigenvalue of $S$. □

## 2.13. Variational Techniques for Image Denoising

**Definition 2.13.1 (Total Variation for Differentiable Functions)**: Let $\Omega \subseteq \mathbb{R}^d$ be an open subset. For $f \in C^1(\Omega, \mathbb{R})$ we define the *total variation* $TV(f)$ of $f$ as

$$TV(f) = \int_\Omega \|\nabla f\|$$

A function is said to be of *bounded variation* if its total variation is finite.

This definition does not work for functions that are *not* differentiable. But we still would like to measure the total variation of nondifferentiable functions. To work towards a definition that works for nondifferentiable functions we start by showing the following lemma.

**Lemma 2.13.1**: Let $\Omega \subseteq \mathbb{R}^d$ be a bounded open subset with a $C^1$ boundary $\partial\Omega$. Let $\varphi \in C_c^1(\Omega, \mathbb{R}^d)$ with $\|\varphi\| \leq 1$ everywhere. Let $f \in C^1\left(\overline{\Omega}, \mathbb{R}\right)$. Then:

$$\int_\Omega f\nabla\cdot\varphi \leq \int_\Omega \|\nabla f\|$$

*Proof*: By the divergence theorem we have

$$\int_\Omega \nabla\cdot(f\varphi) = \int_{\partial\Omega} (f\varphi)\cdot n = 0$$

The r.h.s here is zero because of the compact support of $\varphi$. The l.h.s can be expanded to

$$\int_\Omega \nabla\cdot(f\varphi) = \int_\Omega (\nabla f)\cdot\varphi + \int_\Omega f\nabla\cdot\varphi$$

Thus

$$\begin{aligned}\int_\Omega f\nabla\cdot\varphi = -\int_\Omega (\nabla f)\cdot\varphi &\leq \left|\int_\Omega (\nabla f)\cdot\varphi\right| \\ &\leq \int_\Omega |(\nabla f)\cdot\varphi| \leq \int_\Omega \|\nabla f\|\|\varphi\| \\ &\leq \int_\Omega \|\nabla f\|\end{aligned}$$

□

This result inspires a new definition that works for nondifferentiable functions

**Definition 2.13.2 (Total Variation)**: Let $\Omega \subseteq \mathbb{R}^d$ be a open subset. Let $f \in \mathbb{L}^1(\Omega, \mathbb{R})$. We define the total variation $V(f)$ of $f$ as

$$TV(f) = \sup_\varphi \left\{ \int_\Omega f\nabla\cdot\varphi \right\}$$

where $\varphi \in C_c^1(\Omega, \mathbb{R}^d)$ with $\|\varphi\| \leq 1$.

Note that this definition does not require that the domain $\Omega$ be a bounded set.

**Definition 2.13.3**: Let $f : \mathbb{R}^d \to \mathbb{R}$ be a noisy image, and $\alpha > 0$. We define a denoised image $p_*$ through the optimization

$$p_* = \underset{p}{\operatorname{argmin}}\, \alpha \int \frac{1}{2}|p-f|^2 + \int \frac{1}{2}\|\nabla p\|^2$$

where $p \in C^2\left(\mathbb{R}^d\right)$.

This popular variational regularization method is a special case of *Tikhonov regularization*. The first term in the objective is called the *data term*. The second term is called the *regularization term*. The regularization term forces the denoised image $p_*$ to not have large *variations*, i.e. to be *smooth*. The parameter $\alpha > 0$ determines the strength of the data-fidelity; a lower $\alpha$ means more regularization.

**Proposition 2.13.1**: The solution $p_*$ of the above optimization satisfies

$$p_* - \frac{1}{\alpha}\Delta p_* = f$$

or

$$p_* = \alpha(\alpha - \Delta)^{-1} f$$

The operator $\alpha - \Delta$ is indeed invertible. Notice that in the frequency domain it corresponds to multiplication with $\alpha + \|\omega\|^2$, which can be nicely inverted due to $\alpha > 0$.

*Proof*: Let $E$ be the objective. At the optimum $p_*$ the first order variations of the objective should vanish. Let us inspect the first order variation.

$$\delta E = \alpha \int (p - f)\delta p + \int \nabla p \cdot \nabla \delta p$$

Using Green's first identity (2) this second term can be rewritten as

$$\int \nabla p \cdot \nabla \delta p = -\int \Delta p \delta p$$

where we have used that the boundary term in Green's first identity vanishes because we assume the variation $\delta p$ to vanish far away. We find

$$\delta E = \int (\alpha(p - f) - \Delta p)\delta p$$

setting this equal to zero, and realizing the equality should hold for all variations $\delta p$ tells us that

$$\alpha(p_* - f) - \Delta p_* = 0$$

which can be rewritten into the desired result. □

**Proposition 2.13.2**:

$$\alpha(\alpha - \Delta)^{-1} f = R_\alpha * f$$

where

$$R_\alpha = \alpha \int_0^\infty G_s e^{-\alpha s}\, \mathrm{d}s$$

is the kernel that arises by taking the Laplace transform of the Gaussian kernel w.r.t. scale $s$.

*Proof*: 🚧 □

**Definition 2.13.4**: Let $f : \mathbb{R}^d \to \mathbb{R}$ be a noisy image, and $\alpha > 0$. We define a denoised image $q_*$ through the optimization

$$q_* = \operatorname*{argmin}_q \alpha \int \frac{1}{2}|q - f|^2 + \int \|\nabla q\|$$

where $q \in C^2(\mathbb{R}^d)$.

**Proposition 2.13.3**: The solution $q_*$ of the above optimization satisfies

$$q_* - \frac{1}{\alpha}\nabla \cdot \left(\frac{\nabla q_*}{\|\nabla q_*\|}\right) = f$$

*Proof*: Let $E$ be the objective. Let us inspect the first order variation.

$$\delta E = \alpha \int (q - f)\delta q + \int \frac{\nabla q}{\|\nabla q\|} \cdot \nabla \delta q$$

this second term can be rewritten as

$$\int \frac{\nabla q}{\|\nabla q\|} \cdot \nabla \delta q = -\int \nabla \cdot \left(\frac{\nabla q}{\|\nabla q\|}\right)\delta p$$

where we have used that the boundary term vanishes because we assume the variation $\delta p$ to vanish far away. We find

$$\delta E = \int \left(\alpha(q - f) - \nabla \cdot \left(\frac{\nabla q}{\|\nabla q\|}\right)\right)\delta p$$

setting this equal to zero, and realizing the equality should hold for all variations $\delta p$, tells us that

$$\alpha(q_* - f) - \nabla \cdot \left(\frac{\nabla q_*}{\|\nabla q_*\|}\right) = 0$$

which can be rewritten into the desired result. □

# 3. Invertible Orientation Scores

## 3.1. Group Representation

**Definition 3.1.1 (Group Representation)**: Let $G$ be a group and $X$ a Banach space. A *group representation* $\rho : G \to B(X)$ is a group homomorphism between the group $G$ and the bounded linear maps $B(X)$, i.e.

- $\rho(g_1)\rho(g_2) = \rho(g_1 g_2)$
- $\rho(e) = I_V$

Equivalently, one can think of a group representation as a *bounded linear group action* $\triangleright : G \times X \to X$ defined by $g \triangleright x = \rho(g)x$, which satisfies

- Linearity: $g \triangleright (ax + by) = a(g \triangleright x) + b(g \triangleright y)$
- Boundedness: $g \triangleright$ bounded linear map

- Group Action: $g_1 \triangleright (g_2 \triangleright v) = (g_1 g_2) \triangleright x$
- Identity: $e \triangleright x = x$

If one is talking about finite-dimensional representations the definition is slightly simpler.

**Definition 3.1.2 (Invariant Subspace)**: Let $\rho : G \to B(X)$ be a group representation. A linear subspace $V \subseteq X$ is called *invariant* under $\rho$ if

$$\rho(g)V = V \quad \text{for all } g \in G.$$

Sometimes the invariance of a subspace is stated as

$$\rho(g)V \subseteq V \quad \text{for all } g \in G.$$

but the above two definitions are in fact equivalent.

**Definition 3.1.3 ((Ir)reducible Representation)**: Let $\rho : G \to B(X)$ be a group representation. A representation is called *reducible* if there exists a topologically closed nontrivial linear subspace $\{0\} \neq W \subsetneq V$ that is invariant under $\rho$.

If a representation is not reducible it is called *irreducible*.

Sometimes the reducibility of a representation is not stated as the existence of a nontrivial invariant linear subspace, but instead stated as the existence of a nontrivial *closed* linear subspace, where with closed we mean closed under the action of the representation. We will refrain from using this terminology becauses it can be confused with topologically closed.

The addition of topologically closed in the above definition is important as we will be working with representations on infinite dimensional vector spaces. If one is working with finite dimensional representations the topologically closed part can be omitted from the definition.

**Definition 3.1.4 (Unitary Group Representation)**: Let $G$ be a group and $H$ a Hilbert space. A *unitary* group representation $\rho : G \to U(H)$ is a group representation where the image of $\rho$ consists of unitary operators.

## 3.2. Wavelet Transform

**Definition 3.2.1 (Boring Wavelet Transform)**: Let $G$ be a group with a left invariant Haar measure $\mu$, $H$ a Hilbert space, $\mathcal{U} : G \to U(H)$ a unitary representation, and an element $\psi \in H$ that we call the *wavelet*.

The *boring wavelet transform*[8] $\mathcal{W}_\psi f : G \to \mathbb{C}$ of a element $f \in H$ with wavelet $\psi$ is defined by

$$(\mathcal{W}_\psi f)(g) = \langle \mathcal{U}_g \psi, f \rangle_H.$$

Now this wavelet transform is fine; it is well defined. It even has to following nice lemma

**Lemma 3.2.1 (Wavelet Transform Intertwining)**: Consideer the wavelet transform $\mathcal{W}_\psi : H \to (G \to \mathbb{C})$ we have

$$\mathcal{W}_\psi \circ \mathcal{U}_g = \mathcal{L}_g \circ \mathcal{W}_\psi$$

*Proof*: Let $f \in H$ and $h \in G$ be some dummy elements.

$$\begin{aligned}(\mathcal{W}_\psi(\mathcal{U}_g f))(h) &= \langle \mathcal{U}_h \psi, \mathcal{U}_g f \rangle = \langle \mathcal{U}_{g^{-1}} \mathcal{U}_h \psi, f \rangle \\ &= \langle \mathcal{U}_{g^{-1}h} \psi, f \rangle = (\mathcal{W}_\psi f)(g^{-1}h) \\ &= (\mathcal{L}_g(\mathcal{W}_\psi f))(h),\end{aligned}$$

where we have used the unitarity of $\mathcal{U}$ in the second equality. □

We are however searching for a nicer property, something the Fourier transform on $H = \mathbb{L}^2(\mathbb{R}^d)$ also has: it is inner product preserving. To achieve this we need to change some things.

We first choose for the wavelet transform to instead map to the Hilbert space $\mathbb{L}^2(G)$. However, now that we have chosen this codomain we need to be careful of the domain: we might map outside of $\mathbb{L}^2(G)$.

If for a $f \in H$ we indeed have that $\mathcal{W}_\psi f \in \mathbb{L}^2(G)$ then $f$ is called *admissable*. The admissable elements can be acted upon by the representation to create more admissable elements.

**Lemma 3.2.2**: If $f$ is admissable then $\mathcal{U}_g f$ is admissable for all $g \in G$.

*Proof*: Let $f$ admissable, and $h \in G$ then

$$\begin{aligned}\|\mathcal{W}_\psi \mathcal{U}_h f\| &= \int_G |\langle \mathcal{U}_g \psi, \mathcal{U}_h f \rangle|^2 \mu(dg) \\ &= \int_G |\langle \mathcal{U}_{h^{-1}g} \psi, f \rangle|^2 \mu(dg) \\ &= \int_G |\langle \mathcal{U}_g \psi, f \rangle|^2 \mu(dg) \\ &= \|\mathcal{W}_\psi f\| < \infty\end{aligned}$$

which shows that $\mathcal{U}_h f$ admissable. Notice that we have used that $\mathcal{U}$ is unitary in the second equality, and the left-invariance of the measure $\mu$ in the third equality. □

If we now further assume that the representation $\mathcal{U}$ is irreducible we get the following very strong result.

[8] Also known as the *analysis operator* or *coefficient operator* in abstract harmonic analysis literature[9].

[9] Führ, *Abstract Harmonic Analysis of Continuous Wavelet Transforms*.

**Lemma 3.2.3**: Let $\mathcal{U} : G \to U(H)$ be irreducible, and a nonzero $f \in H$, then the span of $\mathcal{U}_g f$ for all $g \in G$ is dense in $H$.

*Proof*: Let $S$ be this span, and let $\overline{S}$ be the topological closure of $S$. That $\overline{S}$ is invariant under the representation $\mathcal{U}$ is obvious. We also know that $\overline{S}$ in not $\{0\}$ as the nonzero $f$ is in the span. This means, using that $\mathcal{U}$ is irreducible, that we must have that $\overline{S} = H$, i.e. the span $S$ is dense in $H$. □

A somewhat natural thing to demand is that the wavelet $\psi$ itself is admissable. So we assume the existence of a nonzero admissable wavelet $\psi$.

We can now create an appropriate dense domain $D$ of admissable elements for our wavelet transform $\mathcal{W}_\psi$ using the span $U_g\psi$ of the admissable wavelet $\psi$.

**Definition 3.2.2 (Better Wavelet Transform)**: Let $G$ be a group with a left invariant Haar measure $\mu$, $H$ a Hilbert space, $\mathcal{U} : G \to U(H)$ a unitary irreducible representation, and a nonzero admissable wavelet $\psi \in H$. We define the dense domain

$$D = \text{span}\{\mathcal{U}_g\psi \mid g \in G\}$$

The *better wavelet transform* $\mathcal{W}_\psi : D \to \mathbb{L}^2(G)$ with wavelet $\psi$ is defined by

$$(\mathcal{W}_\psi f)(g) = \langle \mathcal{U}_g\psi, f\rangle_H.$$

Notice that we needed the previous two lemmas for this definition to be correct.

We will now show why this adapted wavelet transform is better.

**Lemma 3.2.4 (Better Wavelet Transform is on a lot of subspaces Norm Preserving)**: The better wavelet transform $\mathcal{W}_\psi : D \to \mathbb{L}^2(G)$ is essentially norm preserving in the sense that for all $f = \mathcal{U}_g\psi$

$$\|\mathcal{W}_\psi f\|^2 = C_\psi \|f\|^2$$

where

$$C_\psi = \frac{\|\mathcal{W}_\psi \psi\|^2}{\|\psi\|^2}$$

*Proof*:

$$\begin{aligned}\|W_\psi \mathcal{U}_g\psi\|^2 &= \|\mathcal{L}_g W_\psi \psi\|^2 = \|W_\psi\psi\|^2 \\ &= C_\psi\|\psi\|^2 = C_\psi\|\mathcal{U}_g\psi\|^2\end{aligned}$$

□

In the original work by Grosmann et al.[10] they "show" that $W_\psi$ is actually norm preserving on the *whole* of $D$. I do not trust their argument, but I found a somewhat satisfactory proof myself. One of the lemmas on which they base their argument is as follows.

**Lemma 3.2.5 (Specific Schur's Lemma)**: Let $\rho : G \to U(H)$ be an unitary irreducible representation of a locally compact group $G$. Let $A : (D \subseteq H) \to H$ be a densely defined self-adjoint operator that intertwines with the representation. Then $A$ is a multiple of the identity.

*Proof*: Suppose, for the sake of contradiction, that $A$ is not a multiple of the identity. Then because $A$ is self-adjoint we can perform a spectral decomposition. Through the spectral decomposition, because $A$ is assumed to be *not* a multiple of the identity, we will find a nontrivial closed subspace of $H$. This subspace will be invariant under the representation $\rho$ because $A$ interwines with it. This means $\rho$ is reducible: a contradiction. Thus $A$ is a multiple of the identity. □

The next lemma can also be found in the more modern reference[11].

**Lemma 3.2.6 (Extended Schur's Lemma)**: Let $\rho_1, \rho_2 : G \to U(H_1), U(H_2)$ be two unitary, strongly continuous representations of a locally compact group $G$, where $\rho_1$ is irreducible. Let $A : (D \subseteq H_1) \to H_2$ be a densely defined closed operator that intertwines with the two representations. Then $A$ is (up to a constant) norm preserving.

*Proof*: Because $A$ is a closed densely defined operator Von Neumann's theorem tells us that $A^*A : D(A^*A) \to H_1$ is a densely defined self-adjoint operator. By applying the previous lemma to this operator we find that $A^*A$ is multiple of the identity. Which then tells us that $A$ is in fact a (up to a constant) norm preserving. □

This extended Schur's lemma can be directly applied to our setting of $W_\psi : D \to \mathbb{L}^2(G)$. Now that we have that $W_\psi$ is a (up to a constant) norm preserving map on a dense subspace $D$, we can deduce the existence of a map $W_\psi : H \to \mathbb{L}^2(G)$ that is in fact (up to a constant) inner product preserving on the whole of $H$. By further restricting the codomain we can even make it surjective.

[10] Grossmann, Morlet, and Paul, "Transforms associated to square integrable group representations. I. General results".

[11] Führ, *Abstract Harmonic Analysis of Continuous Wavelet Transforms*.

**Proposition 3.2.1 (Ultimate Wavelet Transform)**: Let $G$ be a group with a left invariant Haar measure $\mu$, $H$ a Hilbert space, $\mathcal{U} : G \to U(H)$ a unitary irreducible representation, $\psi \in H$ a nonzero admissable wavelet, and $D \subseteq H$ the dense domain generated with $\psi$.

There exists a bijective, (up to a constant) inner product preserving wavelet transform $\mathcal{W}_\psi : H \to \big(\mathcal{W}_\psi(H) \subseteq \mathbb{L}^2(G)\big)$, which for $f \in D$ can be expressed as

$$(\mathcal{W}_\psi f)(g) = \langle \mathcal{U}_g \psi, f \rangle_H$$

The up to a constant inner product preservation is expressed as for *all* $f_1, f_2 \in H$ we have

$$\langle \mathcal{W}_\psi f_1, \mathcal{W}_\psi f_2 \rangle_{\mathbb{L}^2(G)} = C_\psi \langle f_1, f_2 \rangle_H$$

## 3.3. Integer Shift Transform

The following transform is not included in the course. I've added it myself to show that the inner product preservation of the wavelet transform can quickly break down even if only **one** of the more subtle assumptions is not satisfied.

**Definition 3.3.1 (Integer Shift Operator)**: Let $n \in \mathbb{Z}$ be an integer shift. We define the *integer shift operator* $\mathcal{U}_n$ by its action on two-sided sequences $f : \mathbb{Z} \to \mathbb{C}$.

$$(\mathcal{U}_n f)(m) = f(m - n)$$

**Definition 3.3.2 (Integer Shift Transform)**: Let $G = (\mathbb{Z}, +)$, $H = l^2(\mathbb{Z})$, and $\mathcal{U} : G \to U(H)$ as the unitary integer shift representation, in the wavelet transform above, we get the *integer shift transform* $\mathcal{W}_\psi : D \to l^2(\mathbb{Z})$:

$$(\mathcal{W}_\psi f)(n) = \sum_{m \in \mathbb{Z}} \overline{\psi(m-n)} f(m)$$

Let use take the wavelet

$$\psi(n) = \begin{cases} 1 & \text{if } n \in \{-1, 0, 1\} \\ 0 & \text{otherwise} \end{cases}$$

This wavelet has norm

$$\|\psi\|^2 = 3$$

and

$$(\mathcal{W}_\psi \psi)(n) = \sum_{m \in \mathbb{Z}} \overline{\psi(m-n)} \psi(m) = \begin{cases} 1 & \text{if } n \in \{-2, 2\} \\ 2 & \text{if } n \in \{-1, 1\} \\ 3 & \text{if } n = 0 \\ 0 & \text{otherwise} \end{cases}$$

So if we valculate the norm of this we get

$$\|\mathcal{W}_\psi \psi\|^2 = 2 \cdot 1^2 + 2 \cdot 2^2 + 3^3 = 37$$

which means $\psi$ is admissable and

$$C_\psi = \frac{\|\mathcal{W}_\psi \psi\|^2}{\|\psi\|^2} = \frac{37}{3}$$

We also have the unitary intertwining relation

$$\mathcal{W}_\psi \circ \mathcal{U}_n = \mathcal{U}_n \circ \mathcal{W}_\psi$$

which means we indeed get the (up to a constant) norm preservation on a lot of subspaces

$$\|\mathcal{W}_\psi \mathcal{U}_n \psi\|^2 = C_\psi \|\mathcal{U}_n \psi\|^2$$

But also note that this wavelet transform is *not* (up to a constant) inner product preserving. Take for example the following $f \in D$

$$f(n) = (\psi - \mathcal{U}_1 \psi)(n) = \begin{cases} 1 & \text{if } n = -1 \\ -1 & \text{if } n = 2 \\ 0 & \text{otherwise} \end{cases}$$

which has norm

$$\|f\|^2 = 1^2 + (-1)^2 = 2$$

We now calculate the transform to be

$$(\mathcal{W}_\psi f)(n) = \sum_{m \in \mathbb{Z}} \overline{\psi(m-n)} f(m) = \begin{cases} 1 & \text{if } n \in \{-2, -1, 0\} \\ -1 & \text{if } n \in \{1, 2, 3\} \\ 0 & \text{otherwise} \end{cases}$$

of which the norm is

$$\|\mathcal{W}_\psi f\|^2 = 3 \cdot 1^2 + 3 \cdot (-1)^2 = 6$$

and we see that

$$\|\mathcal{W}_\psi f\|^2 \neq C_\psi \|f\|^2.$$

This means that $\mathcal{W}_\psi$ can never be (up to a constant) inner product preserving because we have no (up to a constant) norm preservation on the whole of $D$.

So... what went wrong? Why is this wavelet transform not behaving as we "expect" it to behave? I think it has to do with the unitary integer shift representation $\mathcal{U} : \mathbb{Z} \to U(l^2(\mathbb{Z}))$ in fact being *reducible*. This representation being reducible means that it gives us no ensurance that the domain $D$, which is spanned by integer shifts of the wavelet $\psi$, is actually dense in $l^2(\mathbb{Z})$. And indeed, the domain spanned by our chosen wavelet $\psi$ can be shown to be not dense. Consider the following sequence $a : \mathbb{Z} \to \mathbb{C}$

$$a(n) = \begin{cases} 2 & \text{if } n \bmod 3 = 0 \\ -1 & \text{otherwise} \end{cases}$$

The following functional $Z : D \to \mathbb{C}$ is well defined on $D$ and *always* returns zero!

$$Z(f) = \sum_{n \in \mathbb{Z}} a(n) \cdot f(n)$$

For example

$$Z(\psi) = -1 \cdot 1 + 2 \cdot 1 - 1 \cdot 1 = 0$$

This shows that our domain $D$ is a very restricted subspace of $l^2(\mathbb{Z})$ and not even close to being dense: it can not even hope to get near $e_0 \in l^2(\mathbb{Z})$ for which we have $Z(e_0) = 2$!

Notice however that if we had chosen $\psi = e_0$ then $D$ would have been dense, but then the integer shift transform reduces to the identity map.

## 3.4. Roto-Translation Transform

**Proposition 3.4.1 (Roto-Translation Operator is Unitary)**: Let $H = \mathbb{L}^2(\mathbb{R}^d)$ and $G = \mathrm{SE}(d)$. The roto-translation operator $\mathcal{U}_g : H \to H$ is unitary.

*Proof*: We first note that

$$\int f = \int \mathcal{U}_g f,$$

and

$$\mathcal{U}_g(f_1 \cdot f_2) = (\mathcal{U}_g f_1) \cdot (\mathcal{U}_g f_2)$$

Let $f_1, f_2 \in H$ be some dummy Hilbert space elements.

$$\begin{aligned} \langle \mathcal{U}_g f_1, \mathcal{U}_g f_2 \rangle &= \int \overline{\mathcal{U}_g f_1} \cdot \mathcal{U}_g f_2 = \int \mathcal{U}_g\left(\overline{f_1} \cdot f_2\right) \\ &= \int \overline{f_1} \cdot f_2 = \langle f_1, f_2 \rangle, \end{aligned}$$

where in the third equality we used the previous lemma. Thus $\mathcal{U}_g$ preserves the inner product. By definition of being a group representation, $\mathcal{U}_g$ is bijective, and thus $U_g$ is a unitary operator. □

**Proposition 3.4.2 (Roto-Translation Operator in Fourier Domain)**: Let $H = \mathbb{L}^2(\mathbb{R}^d)$ and $\mathcal{U}_{v,R} : H \to H$ the roto-translation operator.

$$(\mathcal{F}\mathcal{U}_{v,R} f)(\omega) = (\mathcal{F} f)(R^{-1}\omega) e^{-i\omega \cdot v}$$

*Proof*: We first show two seperate facts.

$$\begin{aligned} (\mathcal{F}\mathcal{U}_{0,R} f)(\omega) &= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} (\mathcal{U}_{0,R} f)(x) e^{-i\omega \cdot x} \, \mathrm{d}x \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} f(R^{-1}x) e^{-i\omega \cdot x} \, \mathrm{d}x \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} f(y) e^{-i\omega \cdot Ry} \, \mathrm{d}y \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} f(y) e^{-iR^T\omega \cdot y} \, \mathrm{d}y \\ &= (\mathcal{F} f)(R^{-1}\omega) \end{aligned}$$

where the substitution is $x = Ry$.

$$\begin{aligned} (\mathcal{F}\mathcal{U}_{v,I} f)(\omega) &= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} (\mathcal{U}_{v,I} f)(x) e^{-i\omega \cdot x} \, \mathrm{d}x \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} f(x - v) e^{-i\omega \cdot x} \, \mathrm{d}x \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} f(y) e^{-i\omega \cdot (y+v)} \, \mathrm{d}y \\ &= e^{-i\omega v} \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} f(y) e^{-i\omega \cdot y} \, \mathrm{d}y \\ &= e^{-i\omega v} (\mathcal{F} f)(\omega) \end{aligned}$$

where the substitution is $x = y + v$. By combining the two facts with

$$\mathcal{U}_{v,R} = \mathcal{U}_{v,I} \circ \mathcal{U}_{0,R}$$

we obtain the result. □

**Proposition 3.4.3 (Roto-Translation Representation is Reducible)**: The representation $\mathcal{U} : \mathrm{SE}(d) \to B(\mathbb{L}^2(\mathbb{R}^d))$ is reducible.

*Proof*: Consider the closed linear subspace $H_\rho \subset \mathbb{L}^2(\mathbb{R}^d)$ of $\rho$-bandlimited functions

$$H_\rho = \left\{ f \in \mathbb{L}^2(\mathbb{R}^d) \mid \operatorname{ess\,sup} \mathcal{F} f \subseteq B_\rho \subset \mathbb{R}^d \right\}$$

where $B_\rho$ is the open ball of radius $\rho > 0$ in $\mathbb{R}^d$. One can show that this space is invariant under the representation $\mathcal{U}$ by inspecting its behaviour in the Fourier domain using the previous proposition. □

**Definition 3.4.1 (Roto-Translation Transform)**: Taking $G = \mathrm{SE}(d)$, $H = \mathbb{L}^2(\mathbb{R}^d)$, and $\mathcal{U} : G \to U(H)$ as the unitary roto-translation representation, in the wavelet transform above we get the *roto-translation transform*:

$$(\mathcal{W}_\psi f)(v, R) = \int_{\mathbb{R}^d} \overline{\psi(R^{-1}(x - v))} f(x) \, \mathrm{d}x.$$

## 3.5. Similarity Transform

**Definition 3.5.1 (Similarity Group)**: Let $v \in \mathbb{R}^d$ be a translation vector, $R \in \mathbb{R}^{d\times d}$ a rotation matrix, and $a \in \mathbb{R}_{>0}$ a scaling factor. We let a $(v, R, a)$ act $\triangleright$ on $\mathbb{R}^d$ in the obvious way:

$$(v, R, a) \triangleright x = aRx + v \tag{32}$$

We first rotate with $R$, scale with $a$, and then translate with $v$.

We define the $d$-dimensional *similarity group* $\mathrm{SIM}(d)$ as the collection of all similarity transformations with the group operation $\cdot$ being

$$\begin{aligned}&(v_2, R_2, a_2) \cdot (v_1, R_1, a_1)\\ &= (a_2 R_2 v_1 + v_2, R_2 R_1, a_2 a_1)\end{aligned} \tag{33}$$

**Lemma 3.5.1 (Coordinates on Two Dimensional Similarity Group)**:

$$(x_2, y_2, \theta_2, a_2) \cdot (x_1, y_1, \theta_1, a_1) = \begin{pmatrix} x_2 + a_2 x_1 \cos(\theta_2) - a_2 y_1 \sin(\theta_2) \\ y_2 + a_2 x_1 \sin(\theta_2) + a_2 y_1 \cos(\theta_2) \\ \theta_2 + \theta_1 \\ a_2 a_1 \end{pmatrix}$$

**Lemma 3.5.2 (Jacobian on Two Dimensional Similarity Group)**: Let $g = (x, y, \theta, a)$, $g_1 = (x_1, y_1, \theta_1, a_1)$, $g_2 = (x_2, y_2, \theta_2, a_2)$ with

$$g = g_2 \cdot g_1$$

The Jacobian is

$$\frac{\partial g}{\partial g_1} = \begin{pmatrix} a_2\cos(\theta_2) & -a_2\sin(\theta_2) & 0 & 0 \\ a_2\sin(\theta_2) & a_2\cos(\theta_2) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & a_2 \end{pmatrix}$$

with determinant

$$\left|\frac{\partial g}{\partial g_1}\right| = a_2^3$$

**Proposition 3.5.1 (Haar Measure on the Two Dimensional Similarity Group)**: Consider the two dimensional similarity group $G = \mathrm{SIM}(2)$ with coordinates $(x, y, \theta, a)$. The left-invariant Haar measure on $\mathrm{SIM}(2)$ is

$$dg = \frac{1}{a^3}\, dx\, dy\, d\theta\, da$$

*Proof*: From the previous lemma we see that performing a left-translation induces a scaling factor of $a^3$. This factor is balanced by the $\frac{1}{a^3}$ factor in the measure. □

**Definition 3.5.2 (Similarity Transformation Operator)**: Let $g = (v, R, a)$ be a similarity-transformation of $\mathbb{R}^d$. We define the *similarity transformation operator* $\mathcal{V}_{v,R,a}$ by its action on scalar functions $f : \mathbb{R}^d \to \mathbb{C}$:

$$(\mathcal{V}_g f)(x) = \frac{1}{a^{\frac{d}{2}}} f\left(\frac{1}{a} R^{-1}(x - v)\right)$$

or equivalently using the action $\triangleright$ of $\mathrm{SIM}(d)$ on $\mathbb{R}^d$:

$$\mathcal{V}_g f = \frac{1}{a^{\frac{d}{2}}} f \circ (g^{-1} \triangleright)$$

**Lemma 3.5.3 (Similarity Transformation Operator is Unitary)**: Let $H = \mathbb{L}^2(\mathbb{R}^d)$. The similary transformation operator $\mathcal{V}_{v,R,a} : H \to H$ is unitary.

*Proof*: Let $g = (v, R, a)$. We first note that

$$\int \frac{1}{a^{\frac{d}{2}}} \mathcal{V}_g f = \int f,$$

and

$$\mathcal{V}_g f_1 \cdot \mathcal{V}_g f_2 = \frac{1}{a^{\frac{d}{2}}} \mathcal{V}_g (f_1 \cdot f_2).$$

Let $f_1, f_2 \in H$ be some dummy Hilbert space elements.

$$\begin{aligned}\langle \mathcal{V}_g f_1, \mathcal{V}_g f_2 \rangle &= \int \overline{\mathcal{V}_g f_1} \cdot \mathcal{V}_g f_2 = \int \frac{1}{a^{\frac{d}{2}}} \mathcal{V}_g\left(\overline{f_1} \cdot f_2\right)\\ &= \int \overline{f_1} \cdot f_2 = \langle f_1, f_2 \rangle,\end{aligned}$$

Thus $\mathcal{V}_g$ preserves the inner product. By definition of being a group representation, $\mathcal{V}_g$ is bijective, and thus $U_g$ is a unitary operator. □

**Proposition 3.5.2 (Similarity Transform Operator in Fourier Domain)**: Let $H = \mathbb{L}^2(\mathbb{R}^d)$ and $\mathcal{V}_{v,R,a} : H \to H$ the similary transformation operator.

$$\left(\mathcal{F}\mathcal{V}_{v,R,a} f\right)(\omega) = a^{\frac{d}{2}} (\mathcal{F} f)(aR^{-1}\omega) e^{-i\omega\cdot v}$$

*Proof*: We first show the following fact.

$$
\begin{aligned}
(\mathcal{F}\mathcal{V}_{0,I,a}f)(\omega) &= \frac{1}{(2\pi)^{\frac{d}{2}}}\int_{\mathbb{R}^d}(\mathcal{V}_{0,I,a}f)(x)e^{-i\omega\cdot x}\,\mathrm{d}x \\
&= \frac{1}{(2\pi)^{\frac{d}{2}}}\int_{\mathbb{R}^d}\frac{1}{a^{\frac{d}{2}}}f\left(\frac{1}{a}x\right)e^{-i\omega\cdot x}\,\mathrm{d}x \\
&= \frac{1}{(2\pi)^{\frac{d}{2}}}\int_{\mathbb{R}^d}\frac{1}{a^{\frac{d}{2}}}f(y)e^{-i\omega\cdot ay}a^d\,\mathrm{d}y \\
&= a^{\frac{d}{2}}\frac{1}{(2\pi)^{\frac{d}{2}}}\int_{\mathbb{R}^d}f(y)e^{-ia\omega\cdot y}\,\mathrm{d}y \\
&= a^{\frac{d}{2}}(\mathcal{F}f)(a\omega)
\end{aligned}
$$

where the substitution is $x = ay$ with $\mathrm{d}x = a^d\,\mathrm{d}y$. By combining this fact with

$$\mathcal{V}_{v,R,a} = \mathcal{U}_{v,R}\circ\mathcal{V}_{0,I,a}$$

and the previous proposition which showed the behaviour of the roto-translation operator in the Fourier domain, we obtain the result. □

**Lemma 3.5.4 (Similarity Transformation Representation is Irreducible)**: The unitary similarity transformation representation $\mathcal{V}:\mathrm{SIM}(d)\to U(\mathbb{L}^2(\mathbb{R}^d))$ is irreducible.

*Proof*: For the sake of getting a contradiction, suppose $\mathcal{V}$ is reducible and let $V\subseteq\mathbb{L}^2(\mathbb{R}^d) = H$ be a nontrivial topologically closed invariant subspace of the representation $\mathcal{V}$.

Let us inspect the orthogonal complement $V^\perp$ of $V$. This orthogonal complement should be nontrivial if $V$ is nontrivial.

A $f\in H$ is in $V^\perp$ if

$$\langle f, v\rangle = 0 \text{ for all } v\in V$$

More specifically, because $V$ is assumed to be nontrivial and invariant, we can choose a *nonzero* representative $0\neq u\in V$ and write

$$\langle f, \mathcal{V}_g u\rangle = 0 \quad \text{for all } g\in\mathrm{SIM}(d)$$

translating this equation to the Fourier domain we get

$$\left\langle \hat{f}, \hat{u}(aR^{-1}\omega)e^{-i\omega\cdot v}\right\rangle = 0 \quad \text{for all } (v,R,a)\in\mathrm{SIM}(d)$$

Expanding this gives

$$\int_{\mathbb{R}^d}\overline{\hat{f}(\omega)}\hat{u}(aR^{-1}\omega)e^{-i\omega\cdot v}\,\mathrm{d}\omega = 0$$

We can interpret this as a Fourier transform of the first part, and if the Fourier transform is identically 0 the original is also identically 0 almost everywhere

$$\overline{\hat{f}(\omega)}\hat{u}(aR^{-1}\omega) = 0 \quad \text{for all } R, a \text{ and almost all } \omega$$

We know that $0\neq u$ so $0\neq\hat{u}$, and so by choosing $R$ and $a$ in a smart way for any $\omega$, the above equation must imply that $\hat{f} = 0$ which then implies that $f = 0$.

So, we have that $V^\perp = \{0\}$. This means that $\overline{V} = (V^\perp)^\perp = H$, which when taken together with the assumption that $V$ is topologically closed, we get $V = H$. This contradicts the assumption that $V$ is nontrivial, and so $\mathcal{V}$ is irreducible. □

**Definition 3.5.3 (Similarity Transform)**: Taking $G = \mathrm{SIM}(d)$, $H = \mathbb{L}^2(\mathbb{R}^d)$, and $\mathcal{V}: G\to U(H)$ as the unitary irreducible similarity representation, in the wavelet transform definition we get the *similarity transform*:

$$(\mathcal{W}_\psi f)(v,R,a) = \int_{\mathbb{R}^d}\overline{\frac{1}{a^{\frac{d}{2}}}\psi\left(\frac{1}{a}R^{-1}(x-v)\right)}f(x)\,\mathrm{d}x$$

## 3.6. Orientation Score Transform

**Definition 3.6.1 (Homogeneous Space of Positions and Orientations)**: The space of *positions and orientations* is defined as $\mathbb{M}^d = \mathbb{R}^d\times\mathbb{S}^{d-1}$[12]. An element of this space is written as $p = (x,n)\in\mathbb{M}^d$, where $x$ denotes the position and $n$ the orientation.

The group of roto-translations $\mathrm{SE}(d)$ acts naturally on this space

$$(v,R)\triangleright(x,n) = (Rx+v, Rn),$$

turning it into a *homogeneous space*.

**Definition 3.6.2 (Orientation Score Transform)**: Consider $G = \mathrm{SE}(d)$, $H = \mathbb{L}^2(\mathbb{R}^d)$, and $\mathcal{U}: G\to U(H)$ the unitary roto-translation representation. Pick a reference position-orientation $p_0 = (0,n_0)\in\mathbb{M}^d$. Let $\psi\in H$ be a wavelet that is invariant under the stabilizers of $p_0$:

$$\mathcal{U}_s\psi = \psi \quad \text{for all } s\in\mathrm{stab}(p_0)$$

We define the *orientation score transform* $\mathcal{W}_\psi f:\mathbb{M}^d\to\mathbb{C}$ of a $f\in H$ completely analogous to the roto-translation transform

$$(\mathcal{W}_\psi f)(v,Rn_0) = \int_{\mathbb{R}^d}\overline{\psi(R^{-1}(x-v))}f(x)\,\mathrm{d}x$$

Which is well-defined because of the invariance of $\psi$.

[12] Do not confuse all superscripts with repeated (cartesian/manifold) product.

**Proposition 3.6.1 (Orientation Score Transform is Well Defined)**: Pick a $p \in \mathbb{M}^d$. Consider two $g_1, g_2$ that map $p_0$ to $p$. We want to show that

$$(\mathcal{W}_\psi f)(g_1 \rhd p_0) = (\mathcal{W}_\psi f)(g_2 \rhd p_0).$$

*Proof*: We have the equality

$$g_1 \rhd p_0 = g_2 \rhd p_0,$$

acting with $g_2^{-1}$ on both sides gives

$$g_2^{-1} g_1 \rhd p_0 = p_0,$$

so $g_2^{-1} g_1 \in \mathrm{stab}(p_0)$. Using this the proof is immediate:

$$\begin{aligned}(\mathcal{W}_\psi f)(g_1 \rhd p_0) &= \left\langle \mathcal{U}_{g_1}\psi, f \right\rangle = \left\langle \mathcal{U}_{g_2}\mathcal{U}_{g_2^{-1}g_1}\psi, f \right\rangle \\ &= \left\langle \mathcal{U}_{g_2}\psi, f \right\rangle = (\mathcal{W}_\psi f)(g_2 \rhd p_0),\end{aligned}$$

where the third equality uses $\mathcal{U}_s \psi = \psi$ for $s \in \mathrm{stab}(p_0)$. □

## 3.7. Reproducing Kernel Hilbert Spaces

**Definition 3.7.1 (Reproducing Kernel Hilbert Space)**: Let $X$ be an arbitrary set and $H$ a Hilbert space of functions $f : X \to \mathbb{C}$ on that space.[13] The pointwise evaluation at $x$ functional $L_x : H \to \mathbb{C}$ is defined by

$$L_x f = f(x)$$

We say $H$ is a *reproducing kernel Hilbert space* (RKHS) if all such pointwise evaluations functionals $L_x$ are bounded operators.

In practice, once you have a Hilbert space with actual down-to-earth functions you usually already have a RHKS. In other words, it is not entirely straightforward to construct a Hilbert space of functions which is not an RKHS.

Since in a RKHS the pointwise evaluation at $x$ functional $L_x : H \to \mathbb{C}$ is by definition bounded, the Riesz representation theorem tells us that there exists a $K_x \in H$ such that

$$f(x) = L_x(f) = \langle K_x, f \rangle$$

This in turn let us define the *reproducing kernel*

**Definition 3.7.2 (Reproducing Kernel)**: Let $H$ be a RKHS on a set $X$. Define $K_x \in H$ such that

$$f(x) = L_x(f) = \langle K_x, f \rangle$$

We define the *reproducing kernel* $K : X \times X \to \mathbb{C}$ by

$$K(x, y) = \langle K_x, K_y \rangle$$

The reproducing kernel has the properties

$$K_x(y) = L_y(K_x) = \langle K_y, K_x \rangle = K(y, x)$$

Note the slightly awkward change in order of $x$ and $y$. Also note that we have the weird equalities

$$\|K_x\|^2 = \langle K_x, K_x \rangle = K_x(x)$$

Every function $f \in H$ in the Hilbert space can be *reproduced* using just the reproducing kernel in the sense that $f(x) = \langle K(\cdot, x), f \rangle$. This somewhat explains the name "reproducing kernel Hilbert space".

**Definition 3.7.3 (Positive Definite Kernel)**: Let $K : X \times X \to \mathbb{C}$ be a function, called a *kernel*, on a nonempty set $X$. This kernel is called *positive definite* if for all $n \in \mathbb{N}$, $x_1, ..., x_n \in X$ and $c_1, ..., c_n \in \mathbb{C}$ we have

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \overline{c_i} c_j K(x_i, x_j) \geq 0$$

**Proposition 3.7.1 (Reproducing Kernel is Symmetric and Positive)**: The reproducing kernel of a RKHS is conjugate symmetric and positive definite.

*Proof*: The proof is straightforward. For the positive definiteness we have

$$\begin{aligned}\sum_{i=1}^{n} \sum_{j=1}^{n} \overline{c_i} c_j K(x_i, x_j) &= \sum_{i=1}^{n} \sum_{j=1}^{n} \overline{c_i} c_j \left\langle K_{x_i}, K_{x_j} \right\rangle \\ &= \left\langle \sum_{i=1}^{n} c_i K_{x_i}, \sum_{i=1}^{n} c_i K_{x_i} \right\rangle \\ &\geq 0.\end{aligned}$$

And for the conjugate symmetry we have

$$K(x, y) = \langle K_x, K_y \rangle = \overline{\langle K_y, K_x \rangle} = \overline{K(y, x)}.$$

□

Now it would be interesting if every conjugate symmetric and positive definite kernel corresponds to the reproducing kernel of a RKHS. But this is in fact true, and the RKHS is even unique.

**Theorem 3.7.1 (Moore–Aronszajn Theorem)**: Suppose $K : X \times X \to \mathbb{C}$ is a conjugate symmetric, positive definite kernel on a nonempty set $X$. Then there is a unique RKHS $H$[14] on $X$ for which $K$ is a reproducing kernel.

*Proof*: The proof has five parts.

1. We first create an *inner product space* $H_0$ using $K$.
2. We create a *Hilbert space* $H$ as the completion of $H_0$.
3. We check if $H$ is indeed a RKHS.

[13] Note that $H$ need not consist of *all* functions $f : X \to \mathbb{C}$, and that the inner product of $H$ need *not* be the $\mathbb{L}^2$ inner product.

[14] In the lecture notes this unique RKHS is denoted by $\mathbb{C}_K^X$.

4. We check if the reproducing kernel of $H$ is $K$.
5. We show that $H$ is unique.

For all $x \in X$ we define $K_x : X \to \mathbb{C}$ by $K_x(y) = K(y, x)$. Now let $H_0 = \operatorname{span}\{K_x \mid x \in X\}$. We define an inner product $\langle \cdot, \cdot \rangle$ on $H_0$ by

$$\langle K_x, K_y \rangle = K(x, y)$$

and extending this linearly to the whole of $H_0$, i.e.:

$$\left\langle \sum_{i=1}^{m} a_i K_{x_i}, \sum_{j=1}^{n} b_j K_{y_j} \right\rangle_{H_0} = \sum_{i=1}^{m} \sum_{j=1}^{n} \overline{a_i} b_j K(x_i, y_j)$$

Note that even though $H_0$ is a space of functions $f : X \to \mathbb{C}$, the inner product *need not correspond* to a $\mathbb{L}^2$ like inner product. The symmetry of this inner product follows immmediately from the symmetry of $K$, and the positive semi-difiniteness from the fact that $K$ is positive definite. That the inner product is *positive definite*, i.e. non-degenerate, is not immediately obvious. At this point $H_0$ is *just* an inner product space.

We define the Hilbert space $H$ as the completion of $H_0$ with respect to the metric corresponding to the inner product on $H_0$. Then $H$ consists of elements $f$ that are Cauchy sequences $f_n$ in $H_0$, where two sequences $f_n, g_n$ are identified as the same element if $f_n - g_n \to 0$. So any $f \in H$ can be written as

$$f = \lim_{n \to \infty} f_n$$

where $f_n \in H_0$ is a Cauchy sequence:

$$\lim_{n \to \infty} \sup_{m > n} \|f_m - f_n\|_{H_0} = 0$$

That means the inner product on $H$ is

$$\langle f, g \rangle_H = \lim_{n \to \infty} \langle f_n, g_n \rangle_{H_0}$$

and so the norm on $H$ is

$$\|f\|_H = \lim_{n \to \infty} \|f_n\|_{H_0}$$

which are all well-defined due the the Cauchy-ness of the sequences.

This is all kinda abstract, but we can make it significantly more concrete by utilizing the construction of $H_0$ and choosing w.l.o.g. to write $f_n$ as

$$f_n = \sum_{i=1}^{n} a_i K_{x_i}$$

This means we can write

$$f = \sum_{i=1}^{\infty} a_i K_{x_i}$$

The Cauchy sequence criterion becomes:

$$\lim_{n \to \infty} \sup_{m > 0} \left\| \sum_{i=n}^{m} a_i K_{x_i} \right\| = 0$$

The inner product becomes

$$\langle f, g \rangle_H = \sum_{i,j=1}^{\infty} \overline{a_i} b_j K(x_i, y_j)$$

And the norm becomes

$$\|f\|_H = \sum_{i,j=1}^{\infty} \overline{a_i} a_j K(x_i, x_j)$$

So great we have a Hilbert space, but is $H$ actually a RKHS? First let us check if we still have some way of defining pointwise evaluation by seeing if the sequence $f_n(x)$ is Cauchy in $\mathbb{C}$

$$\begin{aligned}(f_m - f_n)(x) &= \sum_{i=n}^{m} a_i K_{x_i}(x) = \sum_{i=n}^{m} a_i K(x, x_i) \\ &= \sum_{i=n}^{m} a_i \langle K_x, K_{x_i} \rangle = \left\langle K_x, \sum_{i=n}^{m} a_i K_{x_i} \right\rangle\end{aligned}$$

Here all equalities are by definition. Now taking norms

$$\begin{aligned}|(f_m - f_n)(x)| &= \left| \left\langle K_x, \sum_{i=n}^{m} a_i K_{x_i} \right\rangle \right| \\ &\leq \|K_x\| \left\| \sum_{i=n}^{m} a_i K_{x_i} \right\|\end{aligned}$$

where the second is the Cauchy inequality. So, indeed, we see $f_n(x)$ inherits the Cauchy-ness of $f_n$, and so $f(x)$ is well-defined and is the limit of $f_n(x)$:

$$f(x) = \sum_{i=1}^{\infty} a_i K_{x_i}(x)$$

That pointwise evaluation is a bounded operator can also be quickly checked.

$$\begin{aligned}|f_n(x)| &= \left| \sum_{i=1}^{n} a_i K_{x_i}(x) \right| = \left| \left\langle K_x, \sum_{i=1}^{n} a_i K_{x_i} \right\rangle \right| \\ &\leq \|K_x\| \left\| \sum_{i=1}^{n} a_i K_{x_i} \right\|\end{aligned}$$

taking the limit $n \to \infty$ implies

$$|f(x)| \leq \|K_x\| \|f\|$$

This means that $H$ is indeed a RKHS.

🚧 Check if the reproducing kernel of $H$ is $K$. Show that $H$ is unique. □

We are still left with an annoying problem: how to calculate the norm of a function $f : X \to \mathbb{C}$ that is in $H$ but of which we don't know yet how the write it in the form $\sum a_i K_{x_i}$? We have the following lemma that let us calculate this norm rather inpractically...

**Lemma 3.7.1**: Let $K : X \times X \to \mathbb{C}$ be a conjugate symmetric, positive definite kernel on a nonempty set $X$, and $H$ the unique corresponding RKHS. Then for all $f \in H$ we have

$$\|f\|^2 = \sup\left\{ \frac{\left|\sum_{i=1}^n b_i f(y_i)\right|^2}{\sum_{i,j=1}^n \overline{b_i} b_i K(y_i, y_j)} \right\}.$$

where $n \in \mathbb{N}$, $b_i \in \mathbb{C}$, and $y_i \in X$.

*Proof*: The Cauchy inequality tells us that:

$$|\langle f, g\rangle|^2 \leq \|f\|^2 \|g\|^2.$$

Shuffling things around we get

$$\|f\|^2 \leq \frac{|\langle f, g\rangle|^2}{\|g\|^2},$$

as long as $g \neq 0$. In fact, one has

$$\|f\|^2 = \sup_{0 \neq g \in H} \frac{|\langle f, g\rangle|^2}{\|g\|^2}.$$

which is directly seen by plugging in $g = f$ if $f \neq 0$, and if $f = 0$ the statement is trivial.

Plugging in the specific form

$$g = \sum_{i=1}^n a_i K_{x_i}$$

for arbitrary $n \in \mathbb{N}$, $b_i \in \mathbb{C}$, and $y_i \in X$, using the defining properties of the RKHS, and that these specifics forms are dense in $H$, we get

$$\|f\|^2 = \sup\left\{ \frac{\left|\sum_{i=1}^n a_i f(x_i)\right|^2}{\sum_{i,j=1}^n \overline{a_i} a_i K(x_i, x_j)} \right\}.$$

□

## 3.8. Space of Bandlimited Square Integrable Functions

**Definition 3.8.1 (Space of Bandlimited Square Integrable Functions)**: Consider the space of $\rho$-bandlimited square integrable in $\mathbb{R}$

$$H = \{f \in \mathbb{L}^2(\mathbb{R}) \mid \operatorname{ess\,sup} \mathcal{F} f \subseteq [-\rho, \rho]\}$$

**Proposition 3.8.1 (RKHS of Bandlimited Square Integrable Continuous Functions)**: The space of $\rho$-bandlimited square integrable in $\mathbb{R}$ is a RKHS.

*Proof*: Let us consider, without loss of generality, $\pi$-bandlimited functions.

It is not immediately obvious why this is even a down-to-earth function space with pointwise evaluation: the definition says we have functions in $\mathbb{L}^2$ for which pointwise evaluation doesn't exist! However, due to the compact essential support of $\mathcal{F} f$ we can in fact point to a *smooth* representative in $C^\infty(\mathbb{R})$ for $f \in \mathbb{L}^2(\mathbb{R})$. We thus, in truth, have square integrable *smooth* functions which are $\rho$-bandlimited.

Due to $\hat{f} = \mathcal{F} f$ being bandlimited we have the inversion formula

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \hat{f}(\omega) e^{i\omega x} \, \mathrm{d}\omega.$$

We can interpret this as

$$f(x) = \frac{1}{\sqrt{2\pi}} \left\langle \hat{f}, e^{i\omega x} 1_{[-\pi,\pi]} \right\rangle_{\mathbb{L}^2(\mathbb{R})}.$$

Applying the Cauchy–Schwarz inequality we get

$$|f(x)| \leq \frac{1}{\sqrt{2\pi}} \left\|\hat{f}\right\|_2 \left\| e^{i\omega x} 1_{[-\pi,\pi]} \right\|_2.$$

Then the Plancherel theorem gives

$$|f(x)| \leq \|f\|_2.$$

So, indeed, pointwise evaluation is a bounded operator. □

As for the reproducing kernel of this space, we first need to know the following facts.

**Lemma 3.8.1 (Sinc, Indicator, and Fourier Transform)**: The inverse Fourier transform of the indicator function $1_{[-a,a]} : \mathbb{R} \to \mathbb{C}$ defined by

$$1_{[-a,a]}(\omega) = \begin{cases} 1 & \text{if } -a \leq \omega \leq a \\ 0 & \text{otherwise} \end{cases}$$

is

$$\mathcal{F}^{-1} 1_{[-a,a]} = \frac{2a}{\sqrt{2\pi}} \operatorname{sinc}(a \cdot)$$

*Proof*: We simply calculate the inverse Fourier transform as

$$\begin{aligned}
&\left(\mathcal{F}^{-1} 1_{[-a,a]}\right)(x) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 1_{[-a,a]}(\omega) e^{i\omega x} \, \mathrm{d}\omega = \frac{1}{\sqrt{2\pi}} \int_{-a}^{a} e^{i\omega x} \, \mathrm{d}\omega \\
&= \frac{1}{\sqrt{2\pi}} \left[ \frac{1}{ix} e^{i\omega x} \right]_{\omega=-a}^{\omega=a} = \frac{1}{\sqrt{2\pi}} \frac{e^{iax} - e^{-iax}}{ix} \\
&= \frac{2}{\sqrt{2\pi}} \frac{\sin(ax)}{x} = \frac{2a}{\sqrt{2\pi}} \operatorname{sinc}(ax)
\end{aligned}$$

with the caveat that the above is not proper when $x = 0$, but one can indeed check that

$$\begin{aligned}\left(\mathcal{F}^{-1}1_{[-a,a]}\right)(0) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} 1_{[-a,a]}(\omega)e^{i\omega 0}\,\mathrm{d}\omega \\ &= \frac{1}{\sqrt{2\pi}}\int_{-a}^{a} 1\,\mathrm{d}\omega \\ &= \frac{2a}{\sqrt{2\pi}}\end{aligned}$$

□

Given the Fourier inversion theorem (for $\mathbb{L}^2(\mathbb{R})$) we then also know that

$$1_{[-a,a]} = \frac{2a}{\sqrt{2\pi}}\mathcal{F}(\mathrm{sinc}(a\,\cdot))$$

One can try to calculate the Fourier transform of sinc directly, but this is awful.

We have the nice special case

$$\left(\mathcal{F}^{-1}1_{[-\pi,\pi]}\right)(x) = \sqrt{2\pi}\,\mathrm{sinc}(\pi x)$$

We can now apply the following "trick"

$$\begin{aligned}f(x) &= \left(\mathcal{F}^{-1}\hat{f}\right)(x) \\ &= \left(\mathcal{F}^{-1}\left(1_{[-\pi,\pi]}\cdot\hat{f}\right)\right)(x) \\ &= \frac{1}{\sqrt{2\pi}}\left(\left(\mathcal{F}^{-1}1_{[-\pi,\pi]}\right) * f\right)(x)\end{aligned}$$

using that $\frac{1}{\sqrt{2\pi}}\mathcal{F}^{-1}1_{[-\pi,\pi]} = \mathrm{sinc}(\pi\,\cdot)$ we see that

$$f(x) = \int_{\mathbb{R}} \mathrm{sinc}(\pi(x-y))f(y)\,\mathrm{d}y$$

which we can intepret as an inner product of the Hilbert space and so the reproducing kernel is

$$K_x(y) = \mathrm{sinc}(\pi(y-x)) = K(y,x)$$

### 3.9. Wavelet Transform on RKHS's

**Definition 3.9.1 (Wavelet Transform to a RKHS)**:

## 4. Tracking in Orientation Scores

### 4.1. Finsler Geometry

**Definition 4.1.1 (Asymmetric Norm)**: Let $V$ be a vector space. A *asymmetric norm* $F : V \to R_{\geq 0}$ is a mapping satisfying

- Positive Homogeneity: $F(\lambda v) = \lambda F(v)$ for all $\lambda \geq 0$
- Triangle Inequality: $F(u+v) \leq F(u) + F(v)$
- Positive Definite: $F(v) = 0 \Rightarrow v = 0$

The only difference between an asymmetric norm and a normal norm is that is is *not* necessarily true that $F(-v) = v$, which explains the name *asymmetric*.

We also sometimes allow an assymmetric norm to return $\infty$. This only works with the conventions that $0 \cdot \infty = 0$, $a + \infty = \infty$ and $a \leq \infty$. These rules are such that we can still make sense of the positive homogeneity, triangle inequality, and positive definiteness.

**Definition 4.1.2 (Finsler Metric & Manifold)**: Let $M$ be a smooth manifold. A *Finsler metric* $\mathcal{F} : TM \to R_{\geq 0}$ is a mapping that is an asymmetric norm at every point $p \in M$, together with a regularity condition.

A *Finsler manifold* is a smooth manifold together with a Finsler metric.

**Definition 4.1.3 (Length of a Curve on a Finsler Manifold)**: Let $M$ be Finsler manifold. We define the *length* $L(\gamma)$ of a continuously differentiable curve $\gamma : [t_0, t_1] \to M$ as

$$L(\gamma) = \int_{t_0}^{t_1} \mathcal{F}(\dot{\gamma}(t))\,\mathrm{d}t$$

**Lemma 4.1.1 (Length of a Curve is Parameterization Invariant)**: Let $M$ be Finsler manifold, $\gamma : [t_0, t_1] \to M$ a continuously differentiable curve, and $\varphi : [s_0, s_1] \to [t_0, t_1]$ a monotonically nondecreasing continiously differentiable function called the *reparameterization*, with $\varphi(s_0) = t_0$ and $\varphi(s_1) = t_1$. We define the *reparameterized* curve $\eta : [s_0, s_1] \to M$ by $\eta = \gamma \circ \varphi$. We have that

$$L(\eta) = L(\gamma)$$

*Proof*: First of we have that

$$\dot{\eta}(s) = \frac{\partial}{\partial s}\eta(s) = \frac{\partial}{\partial s}\gamma(\varphi(s)) = \dot{\gamma}(\varphi(s))\varphi'(s)$$

with $\varphi'(s) \geq 0$ because $\varphi$ is monotonically nondecreasing Then

$$\begin{aligned}L(\eta) &= \int_{s_0}^{s_1} \mathcal{F}(\dot{\eta}(s))\,\mathrm{d}s = \int_{s_0}^{s_1} \mathcal{F}(\dot{\gamma}(\varphi(s))\varphi'(s))\,\mathrm{d}s \\ &= \int_{s_0}^{s_1} \mathcal{F}(\dot{\gamma}(\varphi(s)))\varphi'(s)\,\mathrm{d}s = \int_{t_0}^{t_1} \mathcal{F}(\dot{\gamma}(t))\,\mathrm{d}t \\ &= L(\gamma)\end{aligned}$$

where in the third equality we used the positive homogeneity of the Finsler metric, and in the fourth equality we performed a substitution $t = \varphi(s)$. □

**Definition 4.1.4 (Distance on Finsler Manifold)**: Let $M$ be Finsler manifold. We define a distance $d : M \times M \to \mathbb{R}_{\geq 0}$ by

$$d(p_1, p_2) = \inf\{L(\gamma) \mid \gamma \in \Gamma_T(p_1, p_2)\}$$

where $\Gamma_T(p_1, p_2)$ is the collection of all continuously differentiable curves $\gamma : [0, T] \to M$ with $\gamma(0) = p_1$ and $\gamma(T) = p_2$

Let us use the shorthand $\Gamma_T = \Gamma_T(p_1, p_2)$ for the moment. Due to the parameterization invariance of the length of curves we can *always* choose to reparameterize a curve such that $T = 1$. This means that in the above definition we are free to restrict ourselves to $T = 1$:

$$d(p_1, p_2) = \inf\{L(\gamma) \mid \gamma \in \Gamma_1\}$$

For exactly the same reason, we can also restrict ourselves to curves with $\mathcal{F}(\dot{\gamma}) \leq 1$ or even $\mathcal{F}(\dot{\gamma}) = 1$. This means we *also* have the equalities

$$\begin{aligned} d(p_1, p_2) &= \inf\{L(\gamma) \mid \gamma \in \Gamma_T, \mathcal{F}(\dot{\gamma}) \leq 1\} \\ &= \inf\{L(\gamma) \mid \gamma \in \Gamma_T, \mathcal{F}(\dot{\gamma}) = 1\} \end{aligned}$$

This last restriction is interesting as when $\mathcal{F}(\dot{\gamma}) = 1$ for a curve $\gamma : [0, T] \to M$ then $L(\gamma) = T$. So we *also* have the equality

$$d(p_1, p_2) = \inf\{T \mid \gamma \in \Gamma_T, \mathcal{F}(\dot{\gamma}) = 1\}$$

If we loosen the restriction $\mathcal{F}(\dot{\gamma}) = 1$ to the inequality $\mathcal{F}(\dot{\gamma}) \leq 1$ we get a strictly larger set of curves over which we optimize. But if $\mathcal{F}(\dot{\gamma}) \leq 1$ then $L(\gamma) \leq T$. So we have the inequalities

$$\begin{aligned} d(p_1, p_2) &= \inf\{T \mid \gamma \in \Gamma_T, \mathcal{F}(\dot{\gamma}) = 1\} \\ &\geq \inf\{T \mid \gamma \in \Gamma_T, \mathcal{F}(\dot{\gamma}) \leq 1\} \\ &\geq \inf\{L(\gamma) \mid \gamma \in \Gamma_T, \mathcal{F}(\dot{\gamma}) \leq 1\} \\ &= d(p_1, p_2) \end{aligned}$$

This means, in fact, that we *also* have the equality

$$d(p_1, p_2) = \inf\{T \mid \gamma \in \Gamma_T, \mathcal{F}(\dot{\gamma}) \leq 1\}$$

**Definition 4.1.5 (Energy of a Curve on a Finsler Manifold)**: Let $M$ be Finsler manifold. We define the *$p$-energy* $E^p(\gamma)$, with $p \geq 1$ of a continuously differentiable curve $\gamma : [t_0, t_1] \to M$ as

$$E^p(\gamma) = \int_{t_0}^{t_1} \mathcal{F}(\dot{\gamma}(t))^p \, \mathrm{d}t$$

**Proposition 4.1.1 (Minimizing Energy Minimizes Distance)**: Let $M$ be Finsler manifold, $p_1, p_2 \in M$ two points on the manifold, $p \geq 1$, and consider the energy minimization

$$\inf\{E^p(\gamma) \mid \gamma \in \Gamma_1(p_1, p_2)\}$$

Note that the end time is $T = 1$. We have that this is equal to

$$\ldots = d(p_1, p_2)^p$$

*Proof*: Let us use the shorthand $\Gamma = \Gamma_1(p_1, p_2)$. On $\mathbb{R}_{\geq 0}$ we have that $x \mapsto x^p$ is a convex function. Consider a $\gamma \in \Gamma$. A direct application of *Jensen's inequality* gives

$$\begin{aligned} E^p(\gamma) &= \int_0^1 \mathcal{F}(\dot{\gamma}(t))^p \, \mathrm{d}t \geq \left( \int_0^1 \mathcal{F}(\dot{\gamma}(t)) \, \mathrm{d}t \right)^p \\ &= L(\gamma)^p \end{aligned}$$

Minimizing this over $\Gamma$ we get

$$\inf_{\gamma \in \Gamma} E^p(\gamma) \geq \inf_{\gamma \in \Gamma} L(\gamma)^p$$

If we now consider a curve $\gamma \in \Gamma$ of constant speed $\mathcal{F}(\dot{\gamma}) = L(\gamma)$ (notice that the speed has to be equal to its length due to $T = 1$.) we get that

$$E^p(\gamma) = \int_0^1 \mathcal{F}(\dot{\gamma}(t))^p \, \mathrm{d}t = \int_0^1 L(\gamma)^p \, \mathrm{d}t = L(\gamma)^p$$

Optimizing over the subset of curves of constant speed can only increase the value of the optimization so we also have that

$$\inf_{\gamma \in \Gamma} E^p(\gamma) \leq \inf_{\substack{\gamma \in \Gamma \\ \mathcal{F}(\dot{\gamma}) = L(\gamma)}} L(\gamma)^p = \inf_{\gamma \in \Gamma} L(\gamma)^p$$

where the last equality holds because the length is parameterization independent. Combining the obtained inequalities we get the *equality*

$$\inf_{\gamma \in \Gamma} E^p(\gamma) = \inf_{\gamma \in \Gamma} L(\gamma)^p$$

Now using the nondecreasing monotonicity of $x \mapsto x^p$ on $\mathbb{R}_{\geq 0}$ we can make the final step

$$\inf_{\gamma \in \Gamma} E^p(\gamma) = \inf_{\gamma \in \Gamma} L(\gamma)^p = \left( \inf_{\gamma \in \Gamma} L(\gamma) \right)^p = d(p_1, p_2)^p$$

□

As a direct corollary, using the nondecreasing monotonicity of $x \mapsto x^{\frac{1}{p}}$ on $\mathbb{R}_{\geq 0}$, we get *another* formula for the distance

$$d(p_1, p_2) = \inf\left\{ \sqrt[p]{E^p(\gamma)} \mid \gamma \in \Gamma_1(p_1, p_2) \right\}$$

## 4.2. Reeds-Shepp Car Model

**Definition 4.2.1 (Reeds-Shepp Car Model)**: Consider the space of positions and orientations $M = \mathbb{M}^d$. Let $p = (x, n) \in M$ be a position and orientation and $\dot{p} = (\dot{x}, \dot{n}) \in T_p M \subseteq \mathbb{R}^d \times \mathbb{R}^d$ a tangent vector at this point. Let $c_1, c_2 : M \to \mathbb{R}_{>0}$ be two continuous positive scalar functions on the manifold. We define the *Reeds-Shepp car model*[15] as the Finsler metric $\mathcal{F} : TM \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ with

$$\mathcal{F}(p, \dot{p})^2 = \begin{cases} c_1^2(p)(\dot{x} \cdot n)^2 + c_2^2(p)\|\dot{n}\|^2 & \text{if } \dot{x} = \lambda n, \lambda \in \mathbb{R} \\ \infty & \text{otherwise} \end{cases}$$

We imagine every point $p \in \mathbb{M}^d$ as determining the positions and orientation of a car. A car can not move sideways, only forwards and backwards, and this is reflected in the Finsler metric. The value $c_1(p)$ tells us how "expensive" it is to move forward/backward at the point $p$, and the value $c_2(p)$ tells us how expensive it is to turn. We define the ratio $\xi = \frac{c_1}{c_2}$ which tells us the relative cost of movement versus turning. The reciprocal $\xi^{-1}$ is called the *stiffness* parameter. When the stiffness parameter is large the geodesics of the Reeds-Shepp car model become straighter, which explains the terminology.

**Definition 4.2.2 (Reeds-Shepp Car Model Without Reverse Gear)**: We define the Reeds-Shepp car model *without reverse gear*[17] as the Finsler metric $\mathcal{F}^+ : T\mathbb{M}^d \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ with

$$\mathcal{F}^+(p, \dot{p})^2 = \begin{cases} c_1^2(p)(\dot{x} \cdot n)^2 + c_2^2(p)\|\dot{n}\|^2 & \text{if } \dot{x} = \lambda n, \lambda \geq 0 \\ \infty & \text{otherwise} \end{cases}$$

**Proposition 4.2.1 (Dual Finsler Metric of Reeds-Shepp Car Model)**: Consider the Reeds-Shepp car model Finsler metric $\mathcal{F} : T\mathbb{M}^d \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ as defined above. Let $p = (x, n) \in M$ be a position and orientation and $\hat{p} = (\hat{x}, \hat{n}) \in T_p^* \mathbb{M}^d$ a cotangent vector at this point. The dual Finsler metric $\mathcal{F}^* : T^*\mathbb{M}^d \to \mathbb{R}$ is

$$\mathcal{F}^*(p, \hat{p})^2 = c_1^{-2}(p)(\hat{x} \cdot n)^2 + c_2^{-2}(p)\|\hat{n}\|^2$$

*Proof*:

$$\begin{aligned} \mathcal{F}^*(p, \hat{p}) &= \sup_{\dot{p}} \frac{\langle \hat{p}, \dot{p} \rangle}{\mathcal{F}(p, \dot{p})} \\ &= \sup_{\dot{x} = \lambda n, \dot{n}} \frac{\hat{x} \cdot (\lambda n) + \hat{n} \cdot \dot{n}}{\sqrt{c_1^2 \lambda^2 n \cdot n + c_2^2 \dot{n} \cdot \dot{n}}} \\ &= \sup_{\lambda, \dot{n}} \frac{\left(\frac{\hat{x} \cdot n}{c_1}, \frac{\hat{n}}{c_2}\right) \cdot (c_1 \lambda, c_2 \dot{n})}{\sqrt{(c_1 \lambda, c_2 \dot{n}) \cdot (c_1 \lambda, c_2 \dot{n})}}, \end{aligned}$$

where $(\cdot, \cdot)$ is the inner product on $\mathbb{R}^{1+d}$. We simply apply the Cauchy schwartz (in)equality on $\mathbb{R}^{1+d}$ and get that

$$(c_1 \lambda, c_2 \dot{n}) = \left(\frac{\hat{x} \cdot n}{c_1}, \frac{\hat{n}}{c_2}\right)$$

maximizes the problem, plugging this back into the problem we get

$$\begin{aligned} \ldots &= \frac{\left\| \left(\frac{\hat{x} \cdot n}{c_1}, \frac{\hat{n}}{c_2}\right) \right\|^2}{\left\| \left(\frac{\hat{x} \cdot n}{c_1}, \frac{\hat{n}}{c_2}\right) \right\|} \\ &= \left\| \left(\frac{\hat{x} \cdot n}{c_1}, \frac{\hat{n}}{c_2}\right) \right\| \\ &= \sqrt{c_1^{-2}(p)(\hat{x} \cdot n)^2 + c_2^{-2}(p)\|\hat{n}\|^2} \end{aligned}$$

□

[15] The original paper[16] actually considers a model in which the car has a minimum turning radius; i.e. the paths have a maximal curvature. This simplifies the analysis/paths tremendously.

[16] Reeds and Shepp, "Optimal Paths for a Car That Goes Both Forwards and Backwards".

[17] This looks a lot like a *Dubins path*[18] with no restriction on the curvature. Just as in the Reeds-Shepp case, the restriction on the curvature simplifies everything greatly.

[18] Wikipedia, "Dubins Path".

# Bibliography


Führ, Hartmut. *Abstract Harmonic Analysis of Continuous Wavelet Transforms*. Springer Berlin Heidelberg, 2005. https://doi.org/10.1007/b104912

Grossmann, A., J. Morlet, and T. Paul. "Transforms associated to square integrable group representations. I. General results". *Journal of Mathematical Physics*, no. 10 (1985): 2473–2479. https://doi.org/10.1063/1.526761

Johnson, Mike James. "Relation between Cox-De-boor Recursion and Convolution (B-Spline Basis)", September 2016. https://mathoverflow.net/questions/249141/relation-between-cox-deboor-recursion-and-convolution-b-spline-basis

Reeds, James Alexander III, and Lawrence A. Shepp. "Optimal Paths for a Car That Goes Both Forwards and Backwards". *PACIFIC JOURNAL OF MATHEMATICS*, 1990. https://doi.org/10.2140/pjm.1990.145.367

Schuller, Frederic. "Differential Structures: The Pivotal Concept of Tangent Vector Spaces - Lec 09 - Frederic Schuller", September 2015. https://www.youtube.com/watch?v=UPGoXBfm6Js

Schuller, Frederic. "Lie Groups and Their Lie Algebras - Lec 13 - Frederic Schuller", September 2015. https://www.youtube.com/watch?v=mJ8ZDdA10GY

Schuller, Frederic. “The Fourier Operator - L18 - Frederic Schuller”, March 2016. https://www.youtube.com/watch?v=M-F70q13Jro

Schuller, Frederic. “Topological Manifolds and Manifold Bundles- Lec 06 - Frederic Schuller”, September 2015. https://www.youtube.com/watch?v=uGEV0Wk0eIk

Wikipedia. “Dubins Path”, n.d. https://en.wikipedia.org/wiki/Dubins_path